ARTEFACT

AI IS ABOUT PEOPLE

Data & Agentic Platforms:

The Enablers of AI Innovation

EBOOK DATA PLATAFORM

WHO WE ARE:

Authorship & Team

A equipe da Autoria é composta por profissionais multidisciplinares com expertise em dados, tecnologia, negócios e inovação. Todo time Artefact trabalha de forma colaborativa para entregar soluções de alto impacto, sempre alinhadas às necessidades de nossos clientes.

Tadeu Passos
Associate Director | Data & Software Engineering

Victor Coimbra
Partner Consulting & Data Engineer

Douglas Fonseca
Senior Data Engineer

Thalita Esser
Senior Data Engineer

Paolo Gozdzink
Marketing Specialist

Lay d'Arc
Graphic Designer

ABOUT AUTHORSHIP

Artefact

Artefact accelerates the adoption of data and Artificial Intelligence to positively impact people and organizations. We offer a wide range of services, from strategy to operations, implementing AI solutions across industries to help companies capture the competitive advantage of data and AI transformation.

Sumário

INTRODUCTION

Data as a Strategic Pillar for Growth

FROM DATA LAKE TO AI, EVERYTHING YOU NEED TO MAKE CONFIDENT DECISIONS.

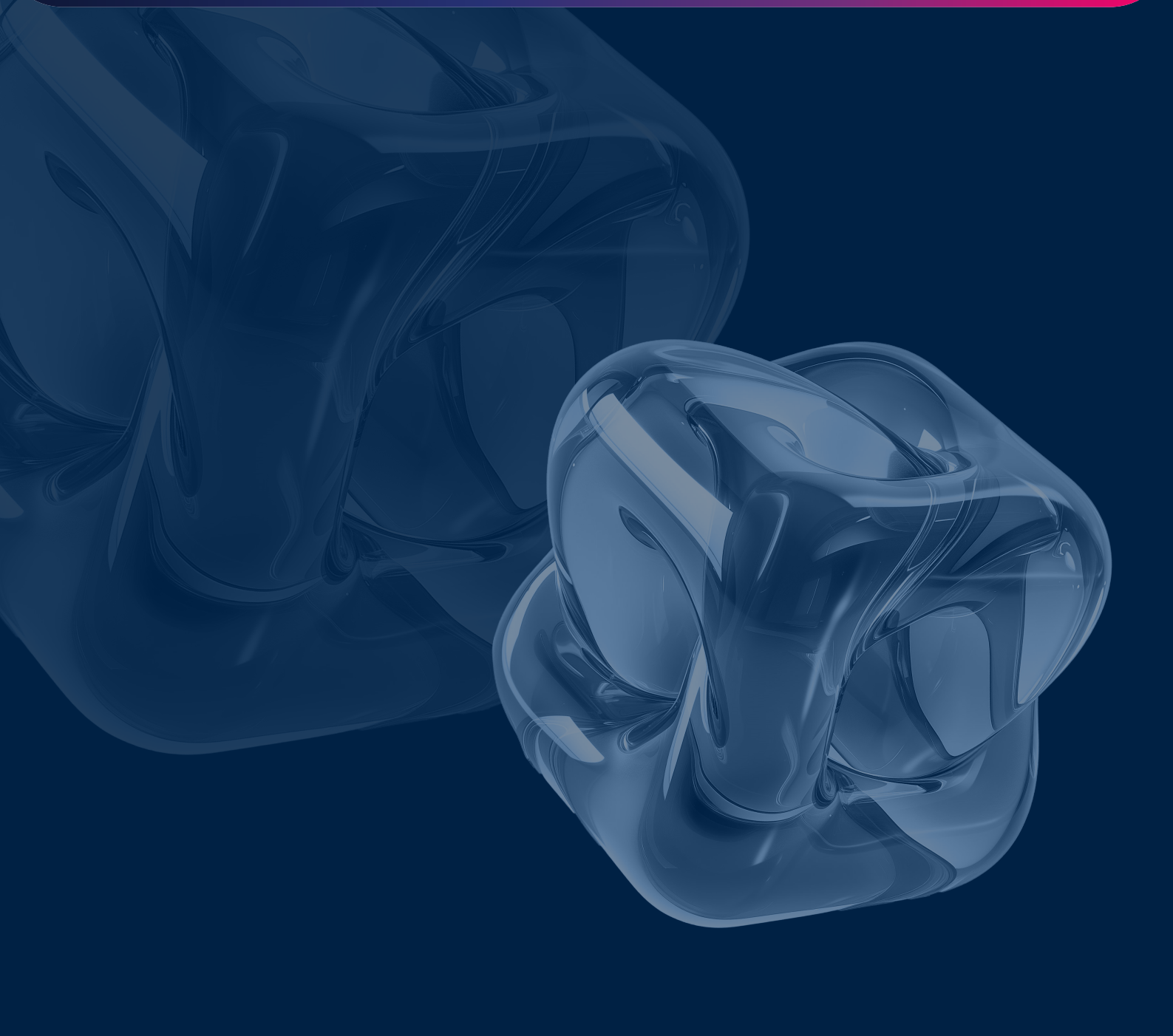
In the following chapters, we'll address various topics related to the growing complexity of today's technological landscape, specifically concerning data platforms. We'll discuss how data architecture has evolved from traditional to modern, with a focus on the reality of the LATAM region, and how this transformation is crucial for overcoming the challenges of the traditional data model.

> We'll deeply explore the revolution of Data Platforms and the New Data Architecture, detailing the foundation-building with Data Lakes and Data Warehouses, and differentiating various types of Data Store layers, such as ODS, Data Lakehouse, Data Mesh, Data Mart, and Data Vault.

Additionally, we'll analyze the fundamental role of data in powering Analytics and Generative Artificial Intelligence (GenAI), the challenges and solutions in implementing Data Platforms, the importance of data governance and security, and present examples of successful transformations. The goal of this e-book is to provide a contribution to better decision-making regarding data architecture and to drive business decision-making.

CHAPTER 1

The Data Architecture Landscape



From Traditional to Modern in the LATAM Region

In recent years, Latin America has undergone a significant transformation in the strategic use of data. The digital acceleration driven by the pandemic, the advance of cloud computing, and, more recently, the emergence of generative artificial intelligence (GenAI) technologies have pressured companies in the region to rethink their data platforms.

While the adoption of Data Platforms is growing, the Latin American landscape still shows a wide variation in terms of maturity, architecture, and the ability to generate value from data. In this context, four key points stand out: the challenges of the traditional data model, the revolution generated by Data Platforms, the so-called new data architecture, and the maturity level of companies in the region.

Challenges of the Traditional Data Model

*For many years, a large portion of companies in Latin America operated in a scenario of **fragmented data, departmental silos, and low standardization.***

Information was scattered across local spreadsheets, legacy systems, and isolated databases, making it difficult to build a unified view of the business. Data exchange between departments was manual or on-demand, and the absence of consistent ingestion, processing, and validation processes compromised analytical reliability.

In this context, strategic decisions were made based on inconsistent or outdated reports, and data governance was practically non-existent—which increased operational risks and reduced the organizations' ability to adapt to market changes.

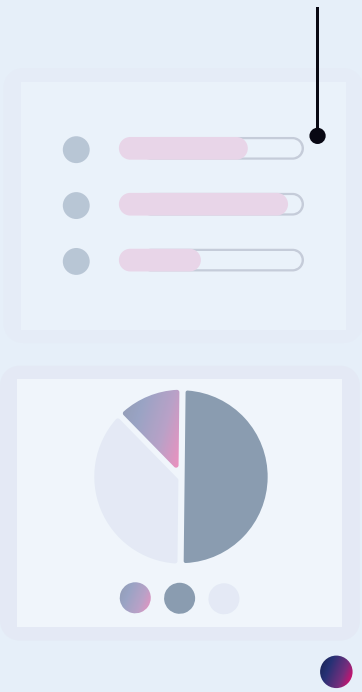
With the advance of cloud technologies, the growth of digitalization, and the pressure for data-driven decisions, it became clear that this decentralized and reactive model no longer met the needs for growth and competitiveness.

> **It was in this scenario that the first initiatives to structure data platforms emerged, with the goal of centralizing, standardizing, and automating the data life cycle.**

This transformation marked the beginning of a new era in organizations, where data began to be treated as a strategic asset—paving the way for what we now call **modern Data Platforms.**

THE DATA PLATFORM REVOLUTION

From the need to centralize and professionalize the use of data in organizations, the first Data Platforms began to emerge, based on data warehouses, ETL pipelines, and BI tools.



The goal was to create a **"single source of truth" that brought together the most relevant business data in a relational**, governed, and secure structure. These environments allowed for the standardization of indicators, automation of extraction routines, and consistent data transformation.

Tools like SAP BW, Oracle, SQL Server, and, more recently, cloud solutions like Snowflake and BigQuery, began to form the heart of corporate analytical operations. With this, business teams could rely on **trustworthy dashboards, periodically updated data, and clearer governance over what each metric represented.**

However, despite the gains in organization, control, and reliability, this centralized and highly technical model began to present challenges as the volume, variety, and velocity of data increased. Data teams became a bottleneck for business areas, which depended on them for any new metric or report. Furthermore, rigid and inflexible ETL processes made it difficult to incorporate new sources and experiment with less structured data. This scenario highlighted the limits of traditional architecture—and opened the way for the emergence of modern, more agile, and scalable platforms.

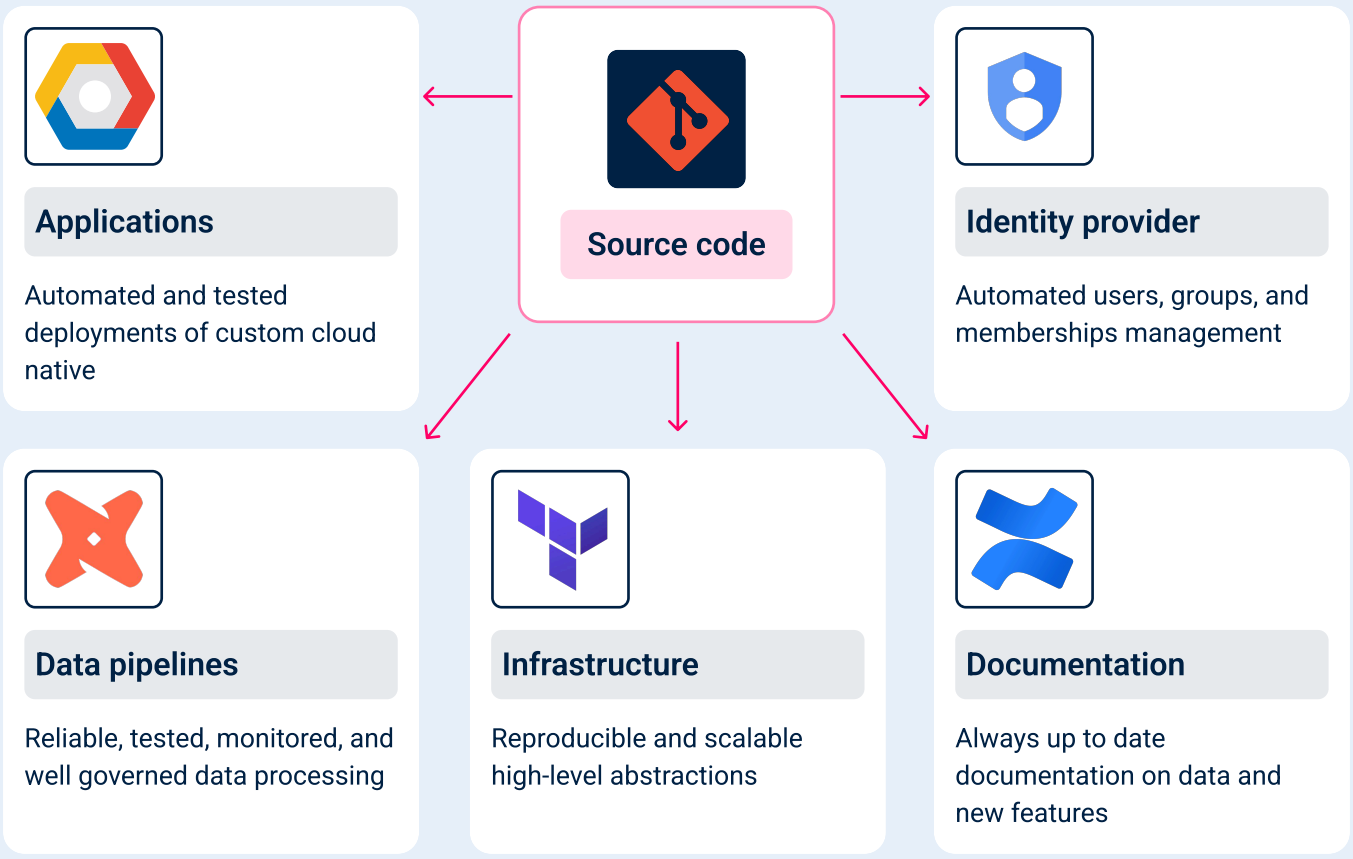
THE NEW DATA ARCHITECTURE

The growing demand for agility, scalability, and autonomy for business areas has driven a new generation of data architectures.

While the traditional model was based on centralization, rigidity, and automation, modern architecture seeks to add governance with flexibility, adopting approaches like Data Lakehouse, Data Mesh, and the use of the Modern Data Stack (with tools like DBT, Airflow, Fivetran, and Snowflake).

In this new paradigm, data is treated as products, with domains responsible for their quality, documentation, and usability, promoting a coordinated decentralization that allows each area of the organization to publish, consume, and explore data with more autonomy and speed. Below you can see an illustration of a modern ecosystem.

Next, we can see an illustration of a modern ecosystem.



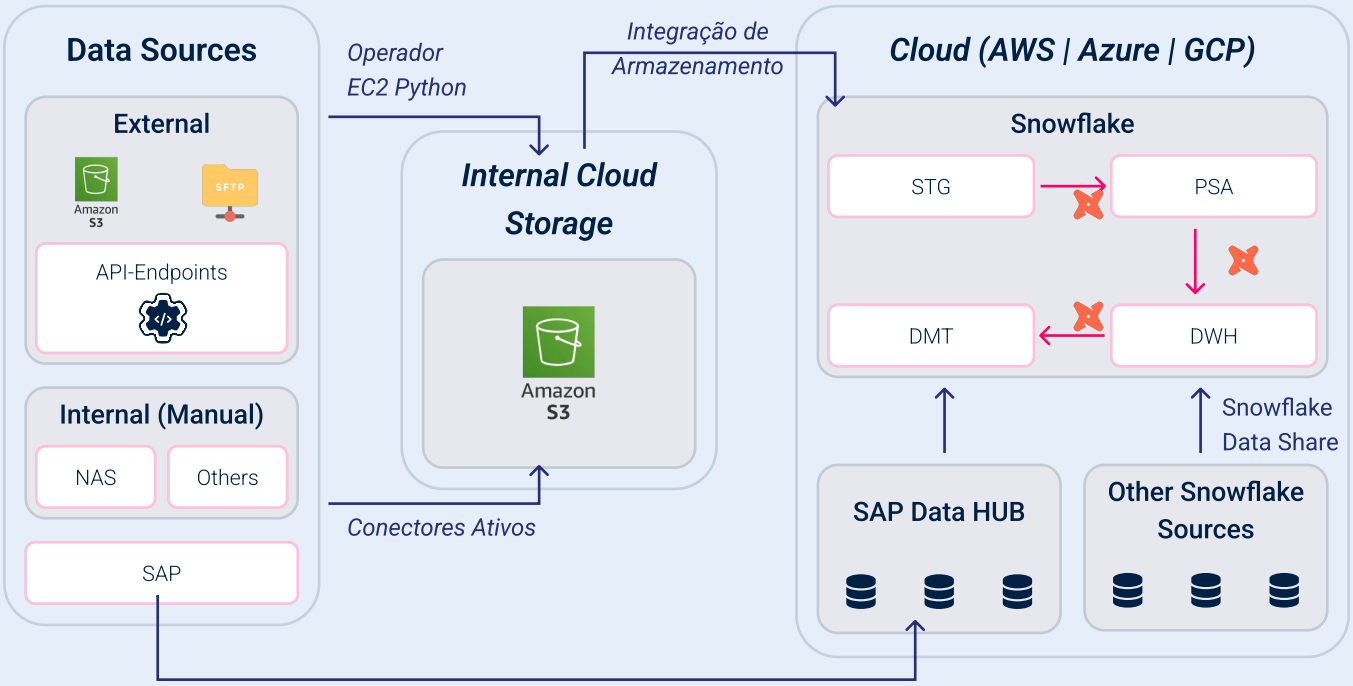
[1] Source: Internal knowledge - Artefact

Additionally, modern architecture relies heavily on the cloud, infrastructure as code, and pipeline automation, which drastically reduces delivery time and facilitates governance at scale. The adoption of well-defined layers* (staging, persistent staging area, Data Warehouse, and Data Mart), versioning practices, continuous monitoring, and intelligent data catalogs creates a more resilient and evolutionary foundation.

> **With this new approach, organizations are better prepared to handle real-time data, a variety of formats, and, especially, to feed machine learning models and GenAI applications, which require reliable, traceable, and ready-to-consume data. In essence, modern architecture transforms the data platform into a strategic enabler of innovation.**

Data Architecture

Illustration of data integration with different sources and formats



COMPANY MATURITY

The maturity of data platforms in Latin American companies varies widely, with many organizations still in the initial stages and others already operating with more advanced structures. Below, we present the five most common levels of maturity, adapted to the reality of the region:

Ad hoc

In this stage, data is treated reactively, without standardized processes or consolidated tools. Each area usually works with its own spreadsheets, local files, and isolated systems, resulting in great duplication of information and a lack of an integrated view. Analysis relies heavily on specific individuals, and there is no data-oriented culture or infrastructure.

Aware

The company begins to recognize the importance of data and realize the negative impacts of the current disorganization. Isolated initiatives for visualizing or centralizing information begin to emerge, usually led by areas like IT, marketing, or finance. There is still no unified strategy, and executive sponsorship is limited, which makes it difficult to advance to more structured stages.

Defined

At this level, the organization implements more formal structures, such as a centralized data warehouse, ETL pipelines, and a dedicated data team. Metrics and indicators are standardized and documented, and BI tools begin to be widely used. However, the model is still centralized and data is not treated as products, limiting the scalability and autonomy of business areas.

Managed

With increasing maturity, the company adopts a modern, cloud-based platform, automated processes, and a modular architecture. Data governance is structured, with clear policies for quality, security, and documentation. Business areas gain autonomy to explore data securely, and strategic decisions are now strongly guided by consistent and reliable insights.

Optimized

In the most advanced stage, the organization operates with real-time data, decentralized architecture, and a high degree of automation. Data is treated as products, with domains responsible for its quality, maintenance, and availability. The company already uses machine learning, AI, and GenAI at scale, with monitored pipelines and robust governance. Below you can see a corporate structure where a data-driven culture is consolidated in all areas, and data is a clear competitive differentiator.

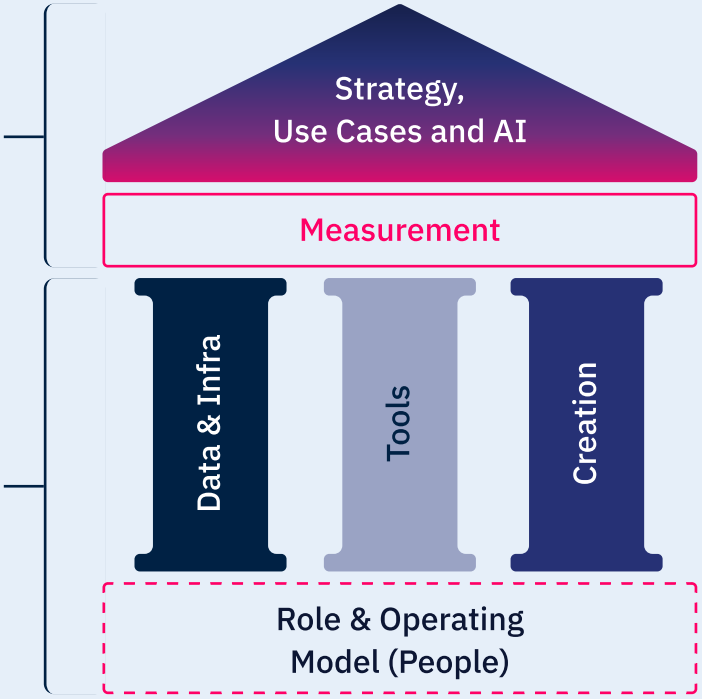
To reach an optimized data state, we need to hit all the key dimensions

Activation

Delivering engaging and data-driven experiences and measuring their performance on all touchpoints.

Enablement

Challenging and recommending tools, data, Role & Operating model to operate data-driven campaigns.



Design relevant digital use cases to improve performance with data, to attract qualified prospects and maximize CLV.

Define a solid measurement scheme in order to continuously improve Marketing efficiency.

Build best-in-class data enablers to collect and process 1P data, and enable 2P data partnerships for enhanced activation.

Implement a consistent & connected tools ecosystem and set up adtech/martech tools used for use cases activation and tracking.

Personalize content and visualisation to compose the best customer journey to drive all the way to conversion.

Define an organizational structure and clear processes adapted to data-driven marketing, by identifying key resources/skills and sharing knowledge with stakeholders.

[1] Fonte: Conhecimento interno - Artefact

CHAPTER 2

Building the Foundation: From Data Lake to Data Warehouse

Differentiating Data Store Layer Types

As companies move forward in the digital age, data architecture becomes increasingly critical for their long-term transformation and success.

> **A modern Data Platform requires an infrastructure that not only stores data efficiently but also allows for agile analysis and solid governance. Choosing Data Store layers is one of the fundamental steps in this process.**
>
> Each type of data storage meets specific needs and is suitable for different levels of maturity and business scenarios.

The Evolution of Data Platforms

Historically, companies relied on traditional **Data Warehouses** to store and analyze structured data. With the increase in data volume and complexity, especially unstructured and semi-structured data, new approaches like **Data Lakes** and **Data Lakehouses** emerged, allowing for greater flexibility. At the same time, the decentralization of data management, proposed by **Data Mesh**, has gained relevance, seeking to solve scalability and agility issues.

> **Digital transformation is forcing companies to rethink their data architecture. The journey of digital maturity directly impacts the choice of storage layers.**

While companies at the beginning of their journey may need flexibility and storage for large volumes of raw data, companies with a higher maturity level generally need more organized and business-oriented solutions.

Understanding the need for each

Data Store type

The different types of Data Store offer unique solutions for specific data challenges.

Companies at different stages of maturity, with different needs for governance, performance, and analysis, can benefit from a combination of layers that best meet their objectives. The choice of which layer to adopt is not just about the technology itself, but also about the type of data the organization needs to manage and how much it is willing to invest in flexibility, governance, and scalability.

Data Lake

A Data Lake is ideal for companies that are starting to accumulate large volumes of data from various sources, such as system logs, sensor data, and social media interactions. It offers the necessary flexibility to handle unstructured and semi-structured data, allowing companies to store everything in a single repository. However, without the proper structure, a Data Lake can quickly become a "data swamp." Companies that are in the initial stage of digital transformation and don't need immediate real-time analysis can benefit from this approach.

Companies that have already reached a level of maturity in their data strategy can benefit from a Data Warehouse to centralize and structure organizational data. With data organized in an optimized format, Data Warehouses are ideal for quick queries and analytical reports, especially serving the needs of Business Intelligence (BI). This model is excellent for companies that need consistent and structured data but can be limited when it comes to working with large volumes of unstructured data.

Data Warehouse

Operational Data Store *(ODS)*

For companies that operate in sectors where quick decisions need to be based on real-time updated data, the ODS is the ideal solution. It acts as an intermediate layer, offering a consolidated view of operational data, with the necessary agility for daily operations. Companies with dynamic business models and constant interactions with customers or transactions, such as e-commerce or financial services, can benefit from this type of layer.

When evolving companies seek to consolidate their structured and unstructured data in a single layer without sacrificing governance and control, the Data Lakehouse is the choice. It integrates the flexibility of a Data Lake with the structure and performance of a Data Warehouse, making it an ideal hybrid solution for organizations that need scalability without losing the ability for quick and efficient analysis.

Data Lakehouse

Data Mesh

Large-scale organizations facing challenges related to governance and scalability can look to the Data Mesh as an alternative to the traditional centralized model. Data Mesh decentralizes data ownership, distributing the responsibility for data management and transformation among different business domains. This approach is particularly useful for large multinational companies or fast-growing companies that need more autonomy in their data teams, without compromising data quality and consistency.

For companies with departments or teams that need quick access to specific data, Data Marts are an efficient solution. They provide highly specialized data subsets for user groups, such as marketing, sales, or finance. Although less robust than Data Warehouses, Data Marts are ideal for organizations that need focused analysis on specific areas without overloading the entire data infrastructure.

Data Mart

Data Vault

The Data Vault is especially relevant for companies that seek a robust approach to data governance and traceability. It provides a flexible and auditable architecture, allowing data to be stored in a scalable and secure way, maintaining the ability to trace every data transformation over time. For companies with high compliance requirements or that operate in regulated sectors, such as health and finance, this architecture offers the necessary level of control.

Connecting Data Store types with the company's journey

The choice of which layer to adopt should consider the company's digital maturity stage, its data objectives, and the complexity of the data it generates.

In an ideal scenario, organizations don't need to limit themselves to a single type of layer. By combining different types of Data Store, they can build a flexible, scalable, and capable data architecture that meets multiple use cases.

As we explore each type of Data Store in the next sections, we'll detail how each one can be applied to transform your data into valuable insights, aligning data architecture with strategic business needs and technological capabilities.

Data Lake: The Foundation of Transformation

A Data Lake is a centralized repository that stores large volumes of raw and unstructured data, such as text files, videos, images, and sensor data. It offers flexibility to store and analyze data from different sources without the need for structuring.

Cloud technologies like Amazon S3, Azure Data Lake Storage, and Google Cloud Storage are commonly used to create and manage Data Lakes, offering scalability and integration with data processing tools like AWS EMR, Azure Databricks, and Google Dataproc.

THE BUSINESS IMPACT OF ADOPTING A DATA LAKE

Companies that adopt Data Lakes have experienced significant improvements in several areas.

Up to 40% reduction in storage costs.

According to a Gartner study, organizations that implement scalable storage technologies like Data Lakes can reduce storage costs by up to 40%, in addition to increasing agility in data integration and access.

These improvements result in faster and more informed decisions, allowing companies to better leverage their information.

Up to 30% increase in data access speed.

Another Forrester Research study revealed that companies that implemented Data Lakes were able to increase data access speed by up to 30%, which boosted business agility and allowed for faster responses to market demands.

This is especially crucial in dynamic sectors, where the speed of decisions can be determinant for maintaining competitiveness.

Use case highlight.

A clear example of success comes from Netflix, which uses a Data Lake to consolidate real-time user behavior data.

The integration of browsing data, viewing history, and feedback allows the company to offer personalized recommendations, which directly contributes to customer satisfaction and user retention.

THE STRATEGIC VALUE OF THE DATA LAKE

The adoption of a Data Lake brings tangible and strategic benefits that directly impact the operation and profitability of companies:

1 Increased Operational Efficiency

Companies that migrated to Data Lakes have reported a significant reduction in time spent on data preparation and integration tasks. This frees up IT and data analysis teams to focus on higher-value tasks, such as extracting insights and innovation. Deloitte found that companies that adopted Data Lakes saw a reduction of up to 40% in operational costs, simplifying the management and integration of data from multiple sources.

2 Real-Time Insights

The Data Lake facilitates real-time data analysis, allowing companies to make quick decisions based on up-to-date information. For example, telecommunications companies like Verizon use Data Lakes to monitor and optimize their infrastructure in real time, ensuring service quality and a better customer experience.

3 Customer Personalization and Innovation

With a Data Lake, it's possible to consolidate data from customer interactions, such as support calls, social media feedback, and sales records, creating a holistic view of consumer behavior. This allows for more personalized marketing campaigns and improvements in customer service. Alibaba and Amazon, for example, use Data Lakes to enhance the customer experience, generating product recommendations based on purchasing behavior and individual preferences.

4 Adoption of Artificial Intelligence and Machine Learning

Data Lakes are essential for implementing artificial intelligence and machine learning technologies, which require large volumes of data to be effective. Companies like General Electric (GE) use Data Lakes to store real-time industrial sensor data, performing predictive maintenance on their machines and preventing costly failures. This improves operational efficiency and reduces costs with corrective maintenance.

CHALLENGES AND CONSIDERATIONS

Although Data Lakes offer great flexibility, they also present challenges, mainly related to governance and data quality. Without well-defined governance, the risk is that the data becomes disorganized, making it difficult to use for effective analysis.

45% of companies face difficulties in managing data quality when implementing Data Lakes, which can compromise the value extracted from this data.

[McKinsey Research]

Therefore, it's essential to invest in data governance practices, such as the use of data catalogs, security policies, and access control. Advanced analytical tools, such as machine learning and Business Intelligence (BI), should also be incorporated to ensure that the value of the data is leveraged to the fullest.

PATHS TO IMPLEMENTING A DATA LAKE

Implementing a Data Lake requires a strategic evaluation and careful execution. The main steps include:

- **Infrastructure Planning:** Choose between cloud or on-premise solutions, considering cost, scalability, and security.
- **Data Governance:** Establish clear policies to ensure the quality, security, and accessibility of stored data.
- **Integration with Analytical Tools:** Integrate the Data Lake with advanced analysis tools, such as machine learning and BI, to extract valuable insights.

Data Warehouse: Centralizing and Structuring Data

A Data Warehouse is a centralized system where structured data from different sources is integrated, organized, and stored for business analysis and report generation. It is designed for quick queries and analysis of large volumes of historical data.

The main cloud technologies for building Data Warehouses include Amazon Redshift, Azure Synapse Analytics, and Google BigQuery, which offer optimized storage solutions and integration with visualization and analysis tools like Power BI and Tableau.

THE STRATEGIC VALUE OF THE DATA WAREHOUSE

The Data Warehouse is designed to centralize data and offer a structure that allows for quick queries and complex analyses.

It is aimed at companies that want to make the most of their historical and operational data, transforming it into actionable information. Unlike the Data Lake, which offers flexibility by storing raw data, the **Data Warehouse organizes and structures information to facilitate trend analysis and the execution of detailed reports.**

The adoption of a Data Warehouse offers a series of advantages, such as:

Quick and Consistent Data Access

By centralizing data from different sources, the Data Warehouse eliminates the complexity of querying scattered data in multiple systems, allowing IT and business teams to access information quickly and accurately.

Improvement in Decision Quality

With structured and organized data, companies can perform more in-depth analyses based on historical data, which improves the quality of strategic decisions. This is particularly valuable for areas like finance, marketing, and operations, which rely on solid data to draw action plans.

Operational Efficiency

The centralization of data reduces redundancy and errors, in addition to providing a clear view of all operations. This results in more efficient processes and greater agility in daily operations.

PRACTICAL CASES OF DATA WAREHOUSE IMPLEMENTATION

Companies that adopt Data Lakes have experienced significant improvements across multiple areas.

1 Financial Sector: Optimization of Regulatory Reports

Large banks and financial institutions, such as Citibank, implement Data Warehouses to centralize transaction, account, and investment data from multiple legacy systems. This facilitates the generation of regulatory reports and compliance with norms, reducing the time and costs involved in the auditing process. The use of a Data Warehouse also improves the accuracy of financial projections and risk analysis.

2 Retail: Sales Analysis and Demand Forecasting

In the retail sector, companies like Walmart use Data Warehouses to consolidate sales, inventory, and customer purchasing behavior data. With this organized information, they can optimize inventory levels and marketing strategies. Furthermore, demand forecasting becomes more accurate, reducing costs with excess or lack of inventory and improving the customer experience.

3 Healthcare: Patient Management and Operational Costs

Hospitals and clinics, such as Mayo Clinic, use Data Warehouses to integrate data from medical records, exams, and treatment history. This data centralization allows healthcare professionals to have a complete view of the patient in one place, which speeds up the diagnosis and treatment process. At the same time, patient data analysis helps identify patterns and optimize operational costs, such as managing medical resources and forecasting demand for care.

4 Telecommunications Industry: Improvement in Customer Experience

Telecommunications companies, such as AT&T, implement Data Warehouses to consolidate data from customers, calls, and service usage. This allows for a precise analysis of consumer behavior, facilitating the personalization of offers and the improvement of customer service. Data centralization also makes the company more efficient in network management and fault forecasting, which improves the quality of the service offered.

COMMERCIAL ADVANTAGES OF THE DATA WAREHOUSE

The implementation of a Data Warehouse offers concrete commercial benefits for companies that seek a structured approach to managing and analyzing data:

Decision-Making Based on Reliable Data

With centralized and organized data, companies can make more accurate and informed decisions. This is especially important in areas like finance, marketing, and operations, where reliable data is essential for drawing strategies and achieving goals.

Reduction of Operational Costs

By centralizing data from different systems, the Data Warehouse eliminates redundancy and improves operational efficiency, resulting in a significant reduction in costs. Companies that adopt this architecture can expect a decrease in time spent on manual processes, reports, and data verification.

Improvement in Governance and Compliance

In highly regulated sectors, such as finance and healthcare, the Data Warehouse is a powerful solution to ensure data governance. It facilitates compliance with norms and regulations, providing a clear audit trail and well-structured information to ensure compliance.

Scalability and Flexibility

The Data Warehouse architecture allows companies to scale their storage capacity as data grows, without compromising performance. Companies can add new data and expand their analytical capabilities as needed, which facilitates adaptation to growth and market changes.

CHALLENGES AND CONSIDERATIONS

Although the Data Warehouse offers countless advantages, it also presents some challenges. One of the main challenges is the ETL process (extraction, transformation, and loading), which can be complex and time-consuming. Furthermore, as data volumes grow, the need for maintenance and optimization of the Data Warehouse increases, which may require continuous investments in technology and specialized personnel.

However, with the right ETL tools and processes and data governance, these challenges can be effectively managed, ensuring that the Data Warehouse continues to offer value to the company in the long term.

IMPLEMENTING A DATA WAREHOUSE

For a successful implementation, companies should follow some key steps:

- **Define Business Objectives:** First and foremost, it's essential to align business objectives with the Data Warehouse architecture, ensuring that it meets the company's specific needs.
- **Choose the Right Infrastructure:** Decide between on-premise or cloud solutions, based on factors like cost, scalability, and security.
- **Implement a Governance Process:** Establish clear data control and quality policies to ensure that the data stored in the Data Warehouse is reliable and accessible.
- **Integration with Analytical Tools:** Ensure that the Data Warehouse is compatible with analysis, reporting, and visualization tools to maximize the value of the stored data.

ODS (Operational Data Store): The Operational View

The Operational Data Store (ODS) is a data repository that integrates real-time operational information from various transactional systems. It provides a consolidated and up-to-date view of operational data, allowing for quick and efficient decisions.

To create a cloud ODS, solutions like Amazon RDS, Azure SQL Database, and Google Cloud SQL can be used, as they offer scalable and high-performance relational databases to store operational data.

WHAT IS THE ODS?

An Operational Data Store (ODS) is a centralized repository that stores real-time operations data. Unlike other storage solutions, the ODS focuses on providing a view of up-to-date and operational data, usually from transactional systems, such as ERP, CRM, or financial systems. This data is integrated, but it is not necessarily transformed or stored as rigorously as in a Data Warehouse. The main objective of the ODS is to ensure that operations teams have access to the most recent data, with the speed and agility necessary to make quick decisions.

HOW THE ODS DRIVES OPERATIONAL EFFICIENCY

For companies that operate in dynamic environments and need a precise view of their operations in real time, the ODS offers considerable value.

It not only improves operational efficiency but also increases the quality of daily decisions, with up-to-date data, ready to be used in real time. The main benefits of an ODS are presented below:

Real-Time Decisions

The main differential of the ODS is to allow operational teams to make immediate decisions based on up-to-date data. This is crucial for sectors like retail, telecommunications, and financial services, where a quick reaction to changes and events is necessary.

Integration of Operational Data

The ODS brings together data from various operational sources and offers a single, consolidated view. This eliminates the need to access different systems to obtain specific information, saving time and resources.

Improvement in Operational Performance

With more accessible and organized operations data, teams can identify bottlenecks, optimize processes, and correct problems quickly. This results in greater efficiency in processes and, consequently, greater customer satisfaction.

PRACTICAL CASES OF ODS IMPLEMENTATION

Companies adopting Data Lakes have seen significant improvements across several areas.

1 Retail: Inventory Management and Customer Service

In the retail sector, companies like Walmart use the ODS to consolidate inventory, sales, and customer interaction data. With an updated view of their inventory and operations, they can make quick decisions, such as product replenishment and dynamic price adjustments, improving efficiency and preventing stockouts. Additionally, real-time data helps with customer service, ensuring that demands are met quickly and effectively.

2 Telecommunications: Network Monitoring and Customer Support

Telecommunications companies, such as AT&T, use ODS to integrate call data, customer records, and network performance in real time. This allows operators to detect service failures or quality drops instantly, in addition to improving the customer experience with more agile support. The ODS helps reduce response time to solve technical and operational problems, ensuring a continuous and high-quality service.

3 Financial Services: Transaction Processing and Risk Management

Banks and insurance companies, such as Banco do Brasil, implement ODS to integrate data from financial transactions, customer history, and risk information. With access to up-to-date data, they can perform real-time analyses to detect fraud, monitor transactions, and make quick decisions related to risk. This improves the ability to respond to fraud and optimizes the credit process.

4 Healthcare: Patient Monitoring and Hospital Management

Hospitals and clinics, such as the Hospital Albert Einstein, use ODS to consolidate patient data, such as medical history, medications, and exams performed. This process allows healthcare professionals to have a complete and updated view of the patient, helping with quick diagnosis and continuous care. The ODS also contributes to the efficient management of hospital resources, such as beds and medical equipment.

COMMERCIAL ADVANTAGES OF THE ODS

*The implementation of an **ODS** can bring concrete commercial benefits to companies that operate in dynamic environments, where decisions need to be quick and based on up-to-date data:*

Increased Agility in Operations

The ODS allows operational teams to have immediate access to up-to-date and integrated data, which accelerates the decision-making process. Companies that use ODS have shown a significant increase in the speed with which they can react to market and daily operational changes.

Reduction of Operational Costs

By consolidating data from various sources into a single repository, the ODS reduces process redundancy and the need to access multiple systems. This contributes to a significant saving of time and resources, resulting in a more efficient operation.

Improvement in Customer Satisfaction

Quick and accurate access to operational data allows companies to respond quickly to customer requests and needs, which, in turn, improves the customer experience. Companies that adopt the ODS can not only solve problems quickly but also anticipate customer needs, creating a competitive differentiator.

CHALLENGES AND CONSIDERATIONS

Although the ODS offers a series of benefits, it also presents challenges, mainly in relation to the complexity of integrating real-time data and the need to ensure that this data is of high quality and consistent. The implementation of an ODS requires good planning to deal with the different data sources and ensure that the platform can deliver accurate and real-time information.

Data governance is another critical point. Although the ODS does not involve a complex transformation of data, it is essential to ensure that the information is updated correctly and that access is controlled efficiently, to avoid operational errors.

HOW TO SUCCESSFULLY IMPLEMENT AN ODS

To ensure that the implementation of an ODS is successful, companies should consider the following steps:

- **Mapping Data Sources:** Identify all operational systems and data sources that will be integrated into the ODS, such as ERP, CRM, financial systems, among others.
- **Choosing the Infrastructure:** Decide between on-premise or cloud solutions, considering the need for scalability, security, and cost.
- **Implementing Data Governance:** Establish clear control and monitoring policies to ensure that the data is accurate, consistent, and accessible to all involved teams.
- **Integration with Operational Tools:** Ensure that the ODS is compatible with the tools used by operational teams, so that the data can be accessed and used efficiently.

Data Lakehouse: The Evolution of the Data Warehouse

The Data Lakehouse combines the flexibility of Data Lakes with the structure of Data Warehouses, allowing the storage and processing of both structured and unstructured data, in a unified way. It provides the best of both worlds, enabling advanced and large-scale analysis.

The main cloud technologies for Data Lakehouses include Databricks on Azure, Amazon Redshift Spectrum, and Google BigLake, which combine the functionalities of Data Lakes with the performance of a Data Warehouse.

WHAT IS A DATA LAKEHOUSE?

The Data Lakehouse is a hybrid architecture that combines the best aspects of a Data Lake and a Data Warehouse.

It allows companies to store data in raw formats, as in the Data Lake, while maintaining the necessary structure for quick queries and analyses, as in a Data Warehouse. This combination allows different types of data—structured, semi-structured, and unstructured—to be managed efficiently in a single platform, offering the flexibility needed to explore data and the essential governance for business analysis.

The concept of Data Lakehouse emerged to overcome the limitations of both separate approaches.

While Data Lakes offer flexibility, they can become difficult to control and use without proper governance. On the other hand, Data Warehouses offer governance, but are less agile when it comes to dealing with unstructured and large-volume data.

HOW THE DATA LAKEHOUSE DRIVES BUSINESS TRANSFORMATION

The adoption of the Data Lakehouse can transform how companies use their data, allowing for a more agile and integrated ecosystem.

It not only solves the challenges faced by traditional Data Warehouse and Data Lake architectures but also offers a series of commercial advantages:

Unified Data View

The Data Lakehouse allows consolidating data from different sources, in various formats, into a single repository. This facilitates the access and analysis of information, which was previously dispersed in multiple systems, optimizing the decision-making process and increasing organizational agility.

Agility and Flexibility for Unstructured Data

Unstructured data, such as videos, images, and texts, are increasingly important for companies. The Data Lakehouse offers the flexibility to store this data while maintaining the ability to perform structured analyses, providing a comprehensive platform for all analytical needs.

Real-Time and Historical Analysis

While the Data Lake facilitates the collection of real-time data, the Data Warehouse is more suitable for historical analysis. The Data Lakehouse combines these two aspects, allowing simultaneous analyses of real-time and historical data, providing a more complete and accurate view of operations and trends.

PRACTICAL EXAMPLES OF DATA LAKEHOUSE IMPLEMENTATION

Organizations adopting Data Lakes have achieved significant improvements in multiple areas.

1 Entertainment Industry: Personalization and User Experience

Streaming companies like Netflix use Data Lakehouses to integrate historical viewing data with real-time data on user behavior. This allows for more precise content recommendations, increasing customer retention and overall user satisfaction. The ability to perform advanced and real-time analyses on large volumes of diversified data makes the Data Lakehouse ideal for personalization and optimization of the customer experience.

2 Retail: Supply Chain Optimization

Retail companies, such as Target, can use a Data Lakehouse to analyze both historical transactional data and real-time data on customer behavior and inventory movement. This allows optimizing the supply chain, forecasting demands, and improving inventory management, ensuring that the right products are available at the right times.

3 Healthcare: Diagnostics and Predictive Forecasts

Hospitals and healthcare organizations like Kaiser Permanente use Data Lakehouses to integrate data from electronic medical records, medical images, and data from wearable devices. With this approach, they can perform predictive and machine learning analyses to forecast health conditions and personalize treatments for patients, improving the quality of care and reducing operational costs.

4 Financial Services: Risk and Compliance Analysis

Financial institutions like JPMorgan Chase use Data Lakehouses to integrate historical financial data with real-time data, such as transactions and market trends. This allows for a more precise analysis of risks and facilitates compliance with financial regulations, while also enabling an agile response to market changes.

COMMERCIAL ADVANTAGES OF THE DATA LAKEHOUSE

The implementation of a Data Lakehouse brings substantial commercial benefits to companies, including:

Greater Operational Efficiency

By centralizing and structuring data from various sources into a single repository, the Data Lakehouse eliminates redundancy and improves the efficiency of data access and use. This reduces maintenance costs and improves response time in daily operations.

Cost Reduction with Storage and Processing

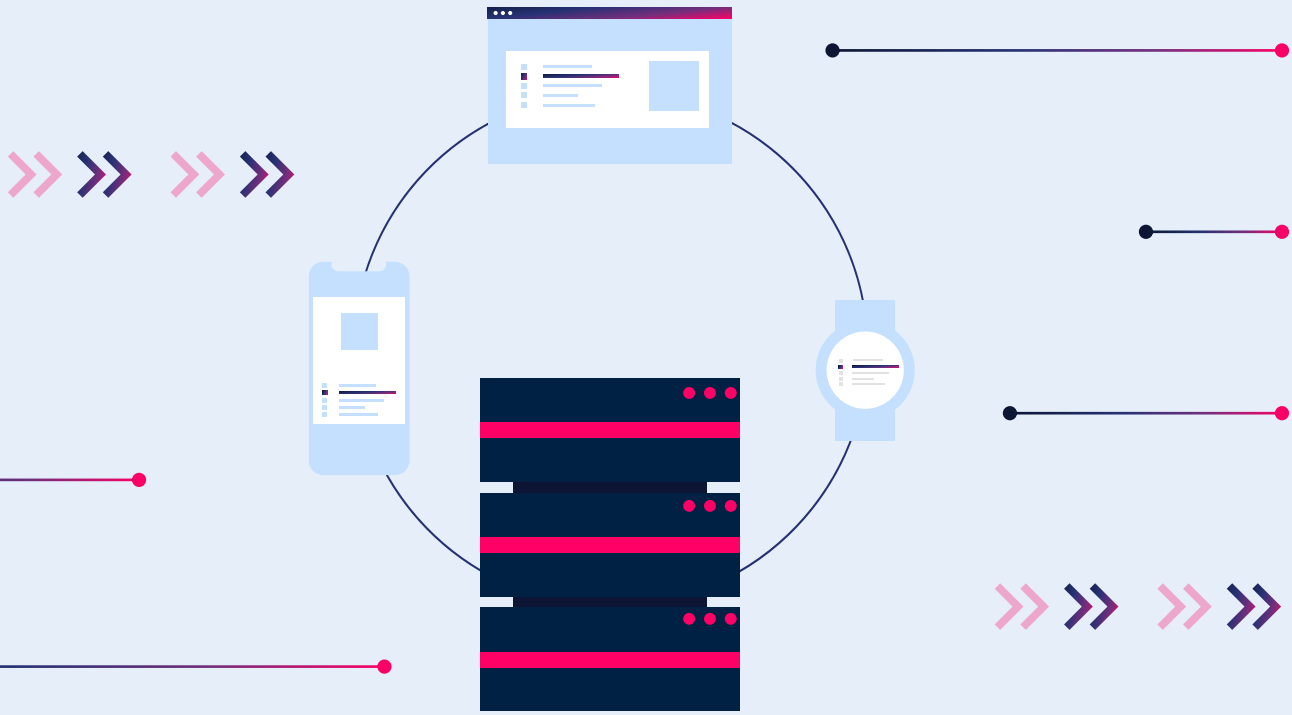
The Data Lakehouse allows companies to store data from different sources and formats more efficiently, without the need to maintain several separate infrastructures. This results in a significant reduction in data storage and processing costs, especially compared to traditional architectures that require separate systems for structured and unstructured data.

Acceleration in Decision-Making

The integration of historical and real-time data into a single repository improves the speed and accuracy of strategic decisions. The Data Lakehouse allows business and IT teams to access the most recent and relevant data to support real-time decisions, increasing agility and responsiveness.

Innovation Capacity with Unstructured Data

The Data Lakehouse is ideal for dealing with unstructured data, such as media, sensor data, and connected device logs. With this capacity, companies can adopt new technologies, such as artificial intelligence and machine learning, to explore data that was previously inaccessible or difficult to analyze.



CHALLENGES AND CONSIDERATIONS

Although the Data Lakehouse offers many benefits, it can also present challenges. Successful implementation requires a good governance strategy to ensure data quality and consistency between historical and real-time data.

Furthermore, companies that already have Data Lakes or Data Warehouses infrastructures may face difficulties in integrating these new architectures, which requires careful planning and a phased approach.

HOW TO IMPLEMENT A DATA LAKEHOUSE

The implementation of a Data Lakehouse requires a strategic and well-planned approach. Some essential steps include:

- **Assessing Data Needs:** Understand what data needs to be integrated and how it will be used to add value to the business.
- **Choosing Infrastructure and Tools:** Decide between on-premise or cloud solutions, and select the appropriate processing and analysis tools.
- **Planning Data Governance:** Establish clear policies for quality, security, and access to ensure that stored data is useful and accessible.
- **Team Training and Capacity Building:** Ensure that IT and business teams know how to explore and take advantage of the Data Lakehouse architecture.

Data Mesh: Decentralizing Data Management

Data Mesh is a decentralized approach to data management, where different business domains manage their own data as "products," allowing for greater autonomy and scalability. It stands out for promoting a distributed governance.

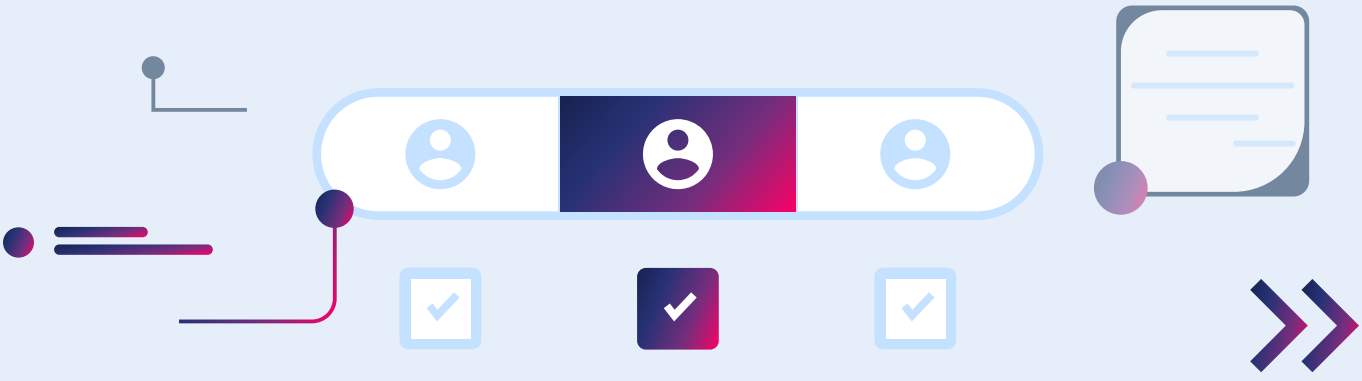
To build a Data Mesh in the cloud, it's common to use tools like Amazon AWS Lake Formation, Azure Data Factory, and Google Cloud Data Fusion, which facilitate the distributed integration and governance of data among different domains.

WHAT IS DATA MESH?

Data Mesh is an architectural concept that proposes the decentralization of data management, distributing the responsibility for data governance and quality among different business domains or areas within an organization.

Instead of a single centralized IT or data team managing all corporate data, each business domain (such as marketing, sales, finance, etc.) becomes responsible for the data it generates, with its own "Data Products," which are organized and accessible data sets for internal and external use.

By adopting Data Mesh, companies can overcome the scalability and governance challenges found in centralized architectures, allowing each business unit to make faster decisions, based on the data it manages directly, while still following governance and interoperability principles between the different domains.



HOW DATA MESH REVOLUTIONIZES DATA GOVERNANCE

Data Mesh offers several advantages that can transform the way companies manage and use their data:

Scalability and Agility

By decentralizing data responsibility, companies can scale their data platforms more easily, as each domain team has the autonomy to manage its data according to its specific needs and requirements. This eliminates the bottlenecks that arise when a single centralized team is responsible for all data.

Autonomy of Business Teams

Each domain has total control over the data it manages, which means that business teams can make faster and more agile decisions. This is particularly useful in dynamic environments, where data needs to be updated and analyzed quickly to meet market demands.

Distributed Governance

Although Data Mesh decentralizes responsibility, it still maintains clear governance principles. The domain teams are responsible for the quality, security, and consistency of the data they manage, but they follow interoperability and access standards established by the organization, ensuring that data can be shared and used effectively among different areas.

PRACTICAL CASES OF DATA MESH IMPLEMENTATION

Businesses that adopt Data Mesh have experienced major improvements across various areas.

1 E-commerce: Personalization and Targeted Marketing

E-commerce companies like Zalando use Data Mesh to decentralize the management of customer, product, and sales data. Each team responsible for user purchasing behavior can optimize its own analyses, with total control over the data generated. This allows for faster and more precise personalization of offers, increasing conversion and customer satisfaction.

2 Financial Services: Risk and Compliance Management

Financial institutions, such as Santander, have implemented Data Mesh to distribute the management of financial and transaction data among different departments (for example, credit, compliance, and risk). Each domain team can manage data relevant to its area, ensuring that decisions about granting credit or risk mitigation strategies are based on up-to-date and well-structured data.

3 Healthcare: Integrating Clinical and Operational Data

Hospitals like the Cleveland Clinic use Data Mesh to integrate clinical and operational data more efficiently. The teams responsible for patient data, such as medical histories and exams, have autonomy to manage this data, while resource management or facility operation teams can access critical information to optimize hospital processes. This not only improves operational agility but also allows healthcare professionals to make more informed and faster decisions.

4 Manufacturing Industry: Production and Quality Management

Manufacturing companies like Siemens are adopting Data Mesh to decentralize the management of production and quality control data. Each plant or production unit can manage its own operational data, optimizing processes locally and at the same time aligning operations with global governance and quality requirements, allowing for more efficient and responsive production.

COMMERCIAL ADVANTAGES OF DATA MESH

Adopting a Data Mesh offers a series of tangible commercial benefits, including:

Greater Speed in Decision-Making

With greater autonomy in business teams, decisions can be made more quickly, without depending on data centralization or a single team to process them. This results in a faster feedback cycle and a more agile ability to adapt to market changes.

Efficiency in Scalability

As an organization grows, the amount of data also increases. Data Mesh allows companies to scale their data operations without overloading a single centralized team, which facilitates continuous growth and adaptation to new requirements.

Reduction of Bottlenecks and Costs

The decentralization of data management eliminates bottlenecks caused by centralization, allowing data teams to focus on the specific needs of their domains, which can reduce operational costs and increase productivity. Furthermore, the reduction of centralized dependencies allows for better use of resources.

Improvement in Governance and Compliance

Data Mesh implements a distributed governance that ensures business teams still follow security and quality guidelines, without compromising autonomy. This facilitates regulatory compliance, as each data domain is responsible for its own governance processes, which facilitates auditing and control.

CHALLENGES AND CONSIDERATIONS

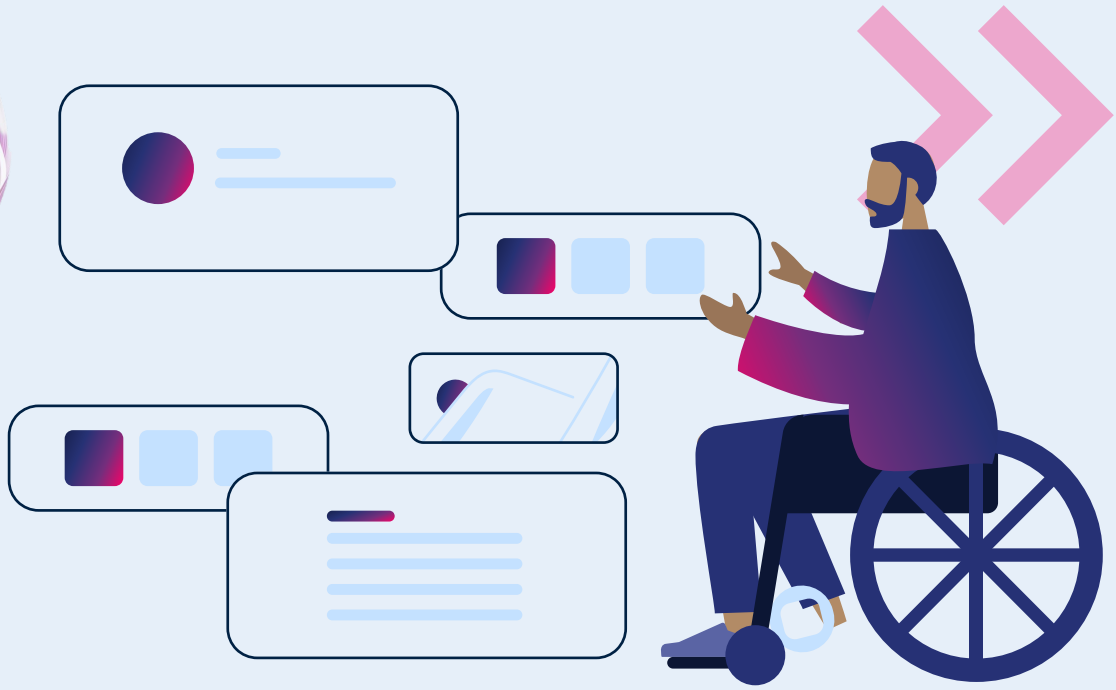
Although Data Mesh brings many benefits, it also presents challenges. The main difficulty is implementing an effective distributed governance across all domains, while also ensuring interoperability between the different data teams. Additionally, companies need to ensure that their domain teams have the necessary skills to manage and govern data properly.

Data Mesh also requires a high level of collaboration among the different data teams and departments, which can be challenging in organizations with rigid hierarchical structures. The integration of legacy systems and adaptation to the new architecture can also represent significant obstacles.

HOW TO IMPLEMENT A DATA MESH

To successfully implement a Data Mesh, companies should follow some key steps:

- **Define Data Domains:** Identify the different business domains (such as marketing, sales, finance) and understand the specific data needs of each one.
- **Establish Governance Standards:** Create clear guidelines for data governance, ensuring that all domains follow standards of quality, security, and interoperability.
- **Invest in Integration and Collaboration Tools:** Implement platforms and tools that allow effective communication between the different domains and ensure that data can be shared efficiently.
- **Empower Data Teams:** Ensure that the teams responsible for data management in each domain have the necessary skills and resources to fulfill their responsibilities for governance and data analysis.



Data Mart: Specific Data for Business Analytics

In many organizations, data is collected and stored in central systems, such as Data Warehouses, but is not always directly applicable to the specific needs of each department or business area. This is where Data Marts come in. A Data Mart is a subset of a Data Warehouse, designed to provide specific and optimized data for specific areas or departments, such as sales, marketing, finance, or operations. By providing data focused on the needs of a specific user group, Data Marts allow for faster analyses and more precise decisions, with less complexity.

Cloud technologies like Amazon Redshift, Azure Synapse Analytics, and Google BigQuery can be used to build Data Marts, leveraging their capabilities for quick queries and scalability for specific departments.

WHAT IS A DATA MART?

A Data Mart is a specialized version of a Data Warehouse, which contains specific data for a function or business area.

While a Data Warehouse stores data from the entire organization in an integrated way, the Data Mart is created to be more focused and simplified, allowing specific teams, such as the marketing or finance team, to access data more relevant to their operations and analyses. This provides greater agility in operations, as the data is structured and optimized for quick and specific analyses.

The Data Mart can be fed by a Data Warehouse, but it can also be a standalone platform in cases where the organization needs a more specific and less complex solution. It allows business areas to obtain detailed insights without relying on complex queries or heavy data processing.

HOW THE DATA MART DRIVES BUSINESS AGILITY

By concentrating the data of a specific area or function, the Data Mart offers a series of direct benefits for companies, including:

Quick and Targeted Analyses

By providing specific data for a business area, the Data Mart allows teams to focus on faster and more targeted analyses, without the need to process large volumes of irrelevant data. This reduces the time needed to obtain actionable insights and increases the productivity of the teams.

Improvement in Decision-Making

Sales, marketing, or finance teams can access data that is directly relevant to their activities, which improves the accuracy of strategic decisions. For example, a marketing team can analyze campaigns more effectively, without the complexity of accessing financial or operational data.

Reduced Load on the Central System

By providing data in a more segmented way, the Data Mart helps to lighten the load on the Data Warehouse, reducing the pressure on the central system and ensuring more efficient performance in daily operations.

PRACTICAL CASES OF DATA MART IMPLEMENTATION

1 Retail: Sales and Customer Behavior Analysis

Large retail chains, such as Best Buy, use Data Marts to provide specific data to their sales and marketing teams. By analyzing segmented data on customer preferences and sales performance, these teams can develop more focused sales strategies, improve promotional campaigns, and personalize offers. The Data Mart also allows for a more agile and quick view of sales trends in different locations.

2 Financial Services: Investment Performance Analysis

Financial institutions like Goldman Sachs use Data Marts to provide specific information on the performance of different assets, investment portfolios, and financial transactions. This allows financial analysts to quickly access the necessary data to assess risk and profitability, without having to process information from other areas of the bank, such as credit operations or accounting.

3 Healthcare: Patient Performance and Treatment Monitoring

Hospitals like the Hospital Israelita Albert Einstein use Data Marts to store patient data related to specific treatments or medical specialties. This allows healthcare professionals to have access to more detailed information about treatments, medical histories, and exam results, facilitating quick decisions regarding patient care and optimization of hospital processes.

4 Manufacturing Industry: Operational Efficiency Analysis

Manufacturing companies like Boeing implement Data Marts to monitor the performance of specific production lines. Each plant or unit can access a Data Mart with real-time operational data, helping to improve production efficiency, reduce waste costs, and quickly identify failures or bottlenecks in the production line.

COMMERCIAL ADVANTAGES OF THE DATA MART

The implementation of a Data Mart can bring direct business benefits in several ways:

Agility and Efficiency in Data Analysis

With more specific and targeted data, teams can access information more quickly and accurately, without the overload of irrelevant data. This results in greater agility in operations and faster decisions, increasing overall efficiency.

Reduction of Infrastructure Costs

As Data Marts are more focused and smaller than Data Warehouses, they can be more economical in terms of data storage and processing. This allows companies to implement more accessible solutions for specific teams, without the need to invest in larger and more complex systems.

Improvement in the Quality of Business Decisions

By providing organized and easily accessible data for business areas, Data Marts help improve the quality of decisions. For example, the marketing team can develop more effective campaigns, based on real-time sales data, while the finance team can make more informed decisions about resource allocation.

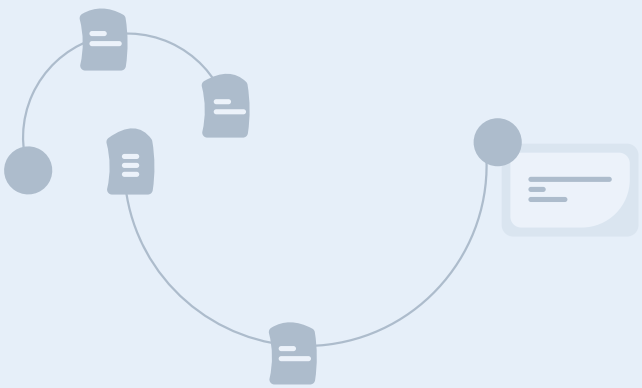
Reduced IT Dependency

With more specific and accessible data directly to business teams, companies can reduce their dependence on IT departments for queries and reports. This frees up IT to focus on more strategic and critical tasks, while business teams gain autonomy to explore and analyze data on their own.

CHALLENGES AND CONSIDERATIONS

Although Data Marts offer a series of benefits, their implementation also presents some challenges. The main consideration is ensuring that the data is updated and consistent between different Data Marts and the central Data Warehouse, if used.

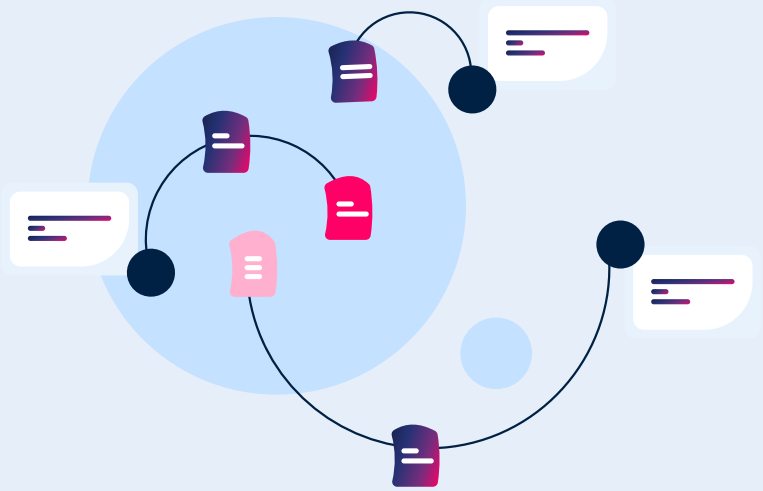
Additionally, the creation of Data Marts in several business areas may require coordination and governance to ensure that data is managed correctly and securely.



HOW TO SUCCESSFULLY IMPLEMENT A DATA MART

To effectively implement a Data Mart, companies should follow some important steps:

- **Define Specific Data Needs:** Each Data Mart must be built based on the specific data needs of a certain department or function. Understanding these requirements is essential to ensure that the Data Mart meets expectations.
- **Establish Data Governance:** Although Data Marts are decentralized, it's important to have central control over data quality and security, ensuring that all teams follow the organization's governance policies.
- **Integration with the Data Warehouse (if necessary):** If the Data Mart is fed by a Data Warehouse, ensuring that the integration between the two systems is done efficiently and without failures is crucial to ensure data consistency.
- **Choose the Right Analysis Tools:** To maximize the value of a Data Mart, it's important to ensure that the data analysis and visualization tools used by business teams are well integrated into the system, allowing for quick and effective queries.



Data Vault: Architecture for Traceability and Governance

The Data Vault is a data architecture that focuses on traceability and governance, allowing data to be stored in a flexible and auditable way. It uses a modular structure of Hubs, Links, and Satellites to organize historical and transactional data.

To implement a Data Vault in the cloud, technologies like Amazon Redshift, Azure Synapse Analytics, and Google BigQuery can be used, leveraging their scalable processing capabilities and the need for robust governance to ensure data integrity.

WHAT IS THE DATA VAULT?

The Data Vault is an architectural approach for data storage that prioritizes traceability, flexibility, and scalability, while ensuring that governance is maintained at all times. It stands out for its ability to handle large-scale data and to integrate data from different sources, including legacy systems and real-time data. Its main characteristic is the construction of a data repository that stores information in a way that any change in the data is easily traced and audited.

*Unlike traditional Data Warehouse approaches, the **Data Vault allows data to be loaded incrementally and easily modified** as new business requirements arise.*

The Data Vault architecture is divided into three main components:

Hubs: Store unique business keys.

Links: Define the relationships between entities.

Satellites: Store the historical and detailed attributes about the keys and relationships.

HOW THE DATA VAULT DRIVES GOVERNANCE AND TRACEABILITY

The adoption of the Data Vault offers significant benefits in terms of data governance and traceability, especially in organizations that need high levels of compliance and control. Among the main benefits are:

Complete Data Traceability

The Data Vault allows any transformation, loading, or modification of data to be traced, creating a detailed audit trail. This is crucial for organizations that operate in regulated sectors or that need a clear understanding of how data has been processed over time.

Flexibility and Scalability

The modular architecture of the Data Vault facilitates adaptation to business changes. New data can be added without affecting existing systems, which ensures that the data platform remains agile and scalable, even as the organization grows or new data sources are integrated.

Simplified Data Governance

The division of data into Hubs, Links, and Satellites makes governance simpler, allowing data teams to efficiently manage data quality and security, without losing the flexibility to integrate new sources. The clarity about where data is stored and how it is related facilitates the implementation of governance policies.

PRACTICAL CASES OF DATA VAULT IMPLEMENTATION

1 Financial Services: Regulatory Compliance and Risk Analyses

Financial institutions, such as HSBC, adopt the Data Vault to ensure the traceability of data in financial transactions and banking operations. This facilitates compliance with regulations such as GDPR (General Data Protection Regulation) and Basel III, as it allows data to be easily audited and any change in historical data to be recorded and traced. Furthermore, the Data Vault facilitates risk analysis by allowing data from different sources to be integrated in a consistent and precise way.

2 Healthcare: Traceability in Patient Data and Medical History

Hospitals and healthcare networks, such as the Cleveland Clinic, use the Data Vault to ensure that patients' medical history is accurately recorded and traced over time. With data coming from different systems (such as electronic medical records, medical device sensors, and laboratory exams), the Data Vault offers a secure and auditable platform to store and integrate this data. This allows healthcare professionals to have complete access to patients' history, facilitating informed clinical decisions and improving patient care.

3 Manufacturing: Production Monitoring and Analysis

Manufacturing companies like General Electric (GE) have adopted the Data Vault to integrate data from industrial sensors, production quality information, and historical maintenance records. With this approach, they can monitor machine performance and predict failures, in addition to ensuring that all data modifications are traced, allowing for a detailed analysis of production efficiency and costs over time.

4 Retail: Purchase and Consumer Behavior Analysis

Retailers like Walmart use the Data Vault to store data from customer transactions, purchasing behavior, and inventory, integrating data from various sources, such as point of sale (POS) systems and e-commerce platforms. This traceability allows marketing teams to create targeted campaigns based on accurate data, while data governance ensures that the data is always up-to-date and aligned with security and privacy policies.

COMMERCIAL ADVANTAGES OF THE DATA VAULT

The implementation of a Data Vault architecture offers a series of commercial advantages for companies, especially in sectors that deal with large volumes of data and require regulatory compliance. The main benefits include:

Rigorous Compliance and Auditing

The Data Vault allows companies to meet regulatory requirements effectively, creating an audit trail that facilitates complete data traceability. This is particularly important in sectors like healthcare, finance, and energy, where compliance is essential and data integrity must be guaranteed.

Scalability and Quick Adaptation

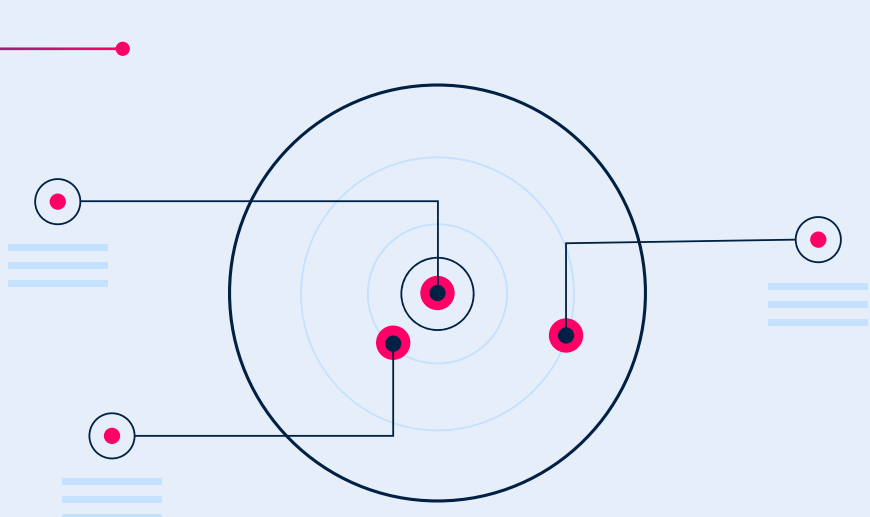
The modular architecture of the Data Vault allows companies to scale their data operations without compromising quality or governance. New data can be integrated quickly, without the need to restructure existing systems, allowing for a more agile adaptation to changes in business or regulation.

Reduction of Operational Risks

Improved traceability and governance provide greater visibility into data, helping companies identify and mitigate risks more quickly. Additionally, data integrity is maintained, which reduces the risks of analysis errors and misguided decisions.

Increased Operational Efficiency

The centralization and organization of data in the Data Vault facilitate the analysis and use of data, reducing the time spent on manual processes and complex queries. This results in greater operational efficiency and allows data teams to focus on higher-value tasks, such as predictive analysis and business intelligence.



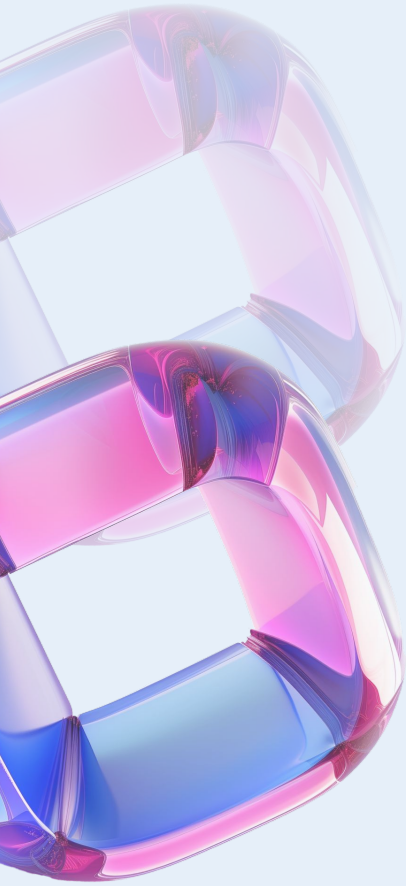
CHALLENGES AND CONSIDERATIONS

Although the Data Vault offers a robust and flexible architecture, its implementation can be more complex than other approaches. The main difficulty is the need for an effective governance strategy to manage data quality and security, especially when there is a large volume and diversity of data sources.

Additionally, the adoption of the Data Vault may require changes in existing data processes and investments in specialized training and tools.

HOW TO IMPLEMENT THE DATA VAULT

To ensure a successful Data Vault implementation, companies should follow some essential steps:



- **Governance Planning:** Define data governance policies that cover all aspects of the data life cycle, from ingestion to analysis.
- **Development of Hubs, Links, and Satellites:** Create a modular architecture that allows data to be stored in an organized and scalable way, with Hubs, Links, and Satellites components.
- **Integration with Legacy Systems and New Data Sources:** Ensure that the Data Vault can integrate data from different sources, including legacy systems and new real-time data, efficiently and without compromising integrity.
- **Team Training:** Invest in the training of IT and data teams to ensure they can use the Data Vault effectively, leveraging all its traceability and governance capabilities.

How to Choose the Best Data Architecture for Your Business

Choosing the right data architecture goes beyond a technical decision. It involves a strategic analysis, taking into account business objectives, budget limitations, and the technical capacity of the team.

The ideal architecture must support the company's long-term vision, while offering the necessary flexibility to adapt to constant changes. Let's explore the main aspects to be considered.

Business Objectives and Specific Needs

If your company needs to deal with large volumes of unstructured data and wants flexibility for future analyses, the most suitable architectures are those that allow data storage in a raw and scalable way, such as Data Lake and Data Lakehouse. Both offer great storage capacity and flexibility, but the Data Lakehouse has the advantage of integrating structured and unstructured data in a single platform, which facilitates more structured analysis, as in a Data Warehouse. If you need quick and consistent reports based on historical structured data, the Data Warehouse is more suitable, as it organizes and optimizes this data for quick queries.

For companies that require agile decisions based on specific data for a department or business area, a Data Mart is ideal. It offers the ability to focus only on what is relevant for each area, allowing teams like sales or marketing to access relevant data with greater efficiency. The Data Mesh stands out when the decentralization of data management is essential, especially in large organizations that operate with multiple data domains. It offers more autonomy, allowing the teams of each data domain to manage their own data sources independently.

BUDGET AND COSTS

In terms of costs, Data Lake and Data Mesh architectures are generally more economical in the initial implementation, as they can be configured with cloud solutions that have lower storage costs. However, as data grows and the need for governance increases, the Data Lake can generate additional costs due to the complexity of managing large volumes of unstructured data, which may require investments in governance and integration tools.

> On the other hand, Data Warehouse and Data Vault require higher investments, especially due to the need for robust infrastructure and ETL (extraction, transformation, and loading) processes. The Data Vault, although more expensive to implement, is ideal for companies that need rigorous control over their data and are in regulated sectors. Data Marts are an intermediate option, as they offer a focused solution, with lower costs, especially when the analysis is limited to a specific area or function, without the complexity of a centralized Data Warehouse.

TECHNICAL CAPACITY OF THE TEAM

If your team is specialized in big data, distributed processing, and large-scale data analysis, architectures like Data Lake and Data Lakehouse stand out, as these solutions offer flexibility to integrate raw data and from various sources. However, they require a well-trained team to deal with the complexity of managing unstructured data.

> If the team is more oriented towards structured data and report analysis, a Data Warehouse or Data Mart solution will be more suitable. These architectures are simpler in terms of integration and maintenance, requiring knowledge in data modeling and SQL. The Data Mesh, in turn, requires a more distributed approach and the ability to coordinate data teams in different domains, which can be challenging in organizations with less mature teams in terms of decentralized data management.

SCALABILITY AND FLEXIBILITY

In terms of scalability, both the Data Lake and the Data Mesh are highly scalable, as they offer the flexibility to integrate new data without rigid structural limitations. The Data Lake is particularly suitable for companies that need to store large volumes of unstructured data, while the Data Mesh is ideal for large and distributed organizations, which need to decentralize responsibility for data to different business units.

> The Data Warehouse, although highly efficient for structured data and optimized for quick queries, can face difficulties when it comes to scaling to handle large volumes of unstructured data or from different sources. The Data Vault, with its modular architecture, also offers scalability, but is more geared towards companies that need an organized and auditable structure, especially when governance is the priority.

IMPLEMENTATION SPEED

When it comes to implementation speed, solutions like Data Lake and Data Mart can be implemented quickly, especially if they are configured on cloud platforms. The Data Mart is particularly fast to implement, as it is focused on a specific data subset and a business area, without the need for a complex data transformation.

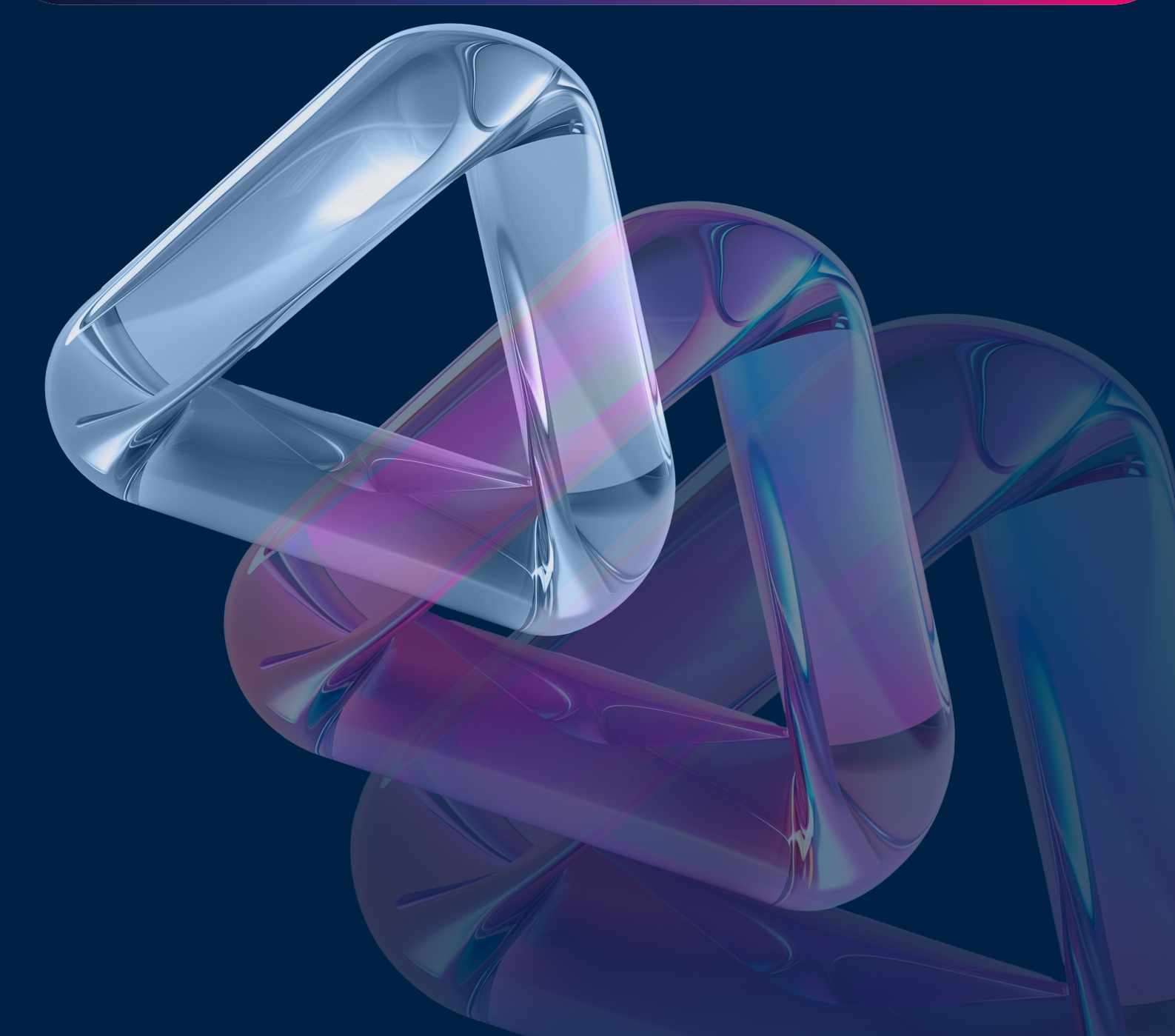
> The Data Warehouse and the Data Vault, due to their complexity and need for detailed data modeling, can take more time to be implemented, but offer greater control and governance in the long term. The Data Mesh, because it is a decentralized architecture, can take more time to be structured and require additional coordination effort among the different data domains.

In summary, choosing the ideal data architecture depends on your business's specific requirements, available budget, and technical capacity. If the goal is to deal with large volumes of unstructured data with flexibility and scalability, a Data Lake or Data Mesh solution may be the most suitable. For companies that need rigorous control over structured data and quick queries, a Data Warehouse or Data Vault may be more appropriate.

On the other hand, for those who are looking for a focused and more economical solution for a specific department, the Data Marts offer a good balance between cost and agility. The important thing is to align the company's data needs with the architectural options, ensuring that the chosen solution is capable of meeting current and future business demands, while also being financially viable and technically executable.

CHAPTER 3

AI & Analytics
Data



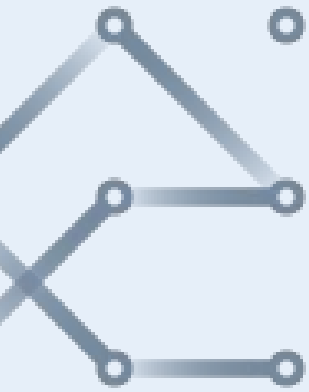
The Role of Data in Powering Analytics and GenAI

The evolution of data platforms is one of the main drivers of technological innovations in organizations. This transformation includes the ability to collect, process, and analyze massive volumes of data in real time, allowing for the development of advanced analytics solutions.

> **While traditional BI tools still play a fundamental role, modern platforms offer a new perspective, with predictive and automation analyses.**

In this chapter, the challenges and strategies involved in implementing modern data platforms that support these solutions are explored. Also, the importance of a solid, scalable data architecture aligned with governance needs is presented. In addition, real examples of how these platforms are being leveraged to create value in the most diverse industries are separated, with the transformation of data into strategic decisions through the combination of GenAI and Advanced Analytics.

How Data Architecture Drives Advanced Analytics Use Cases

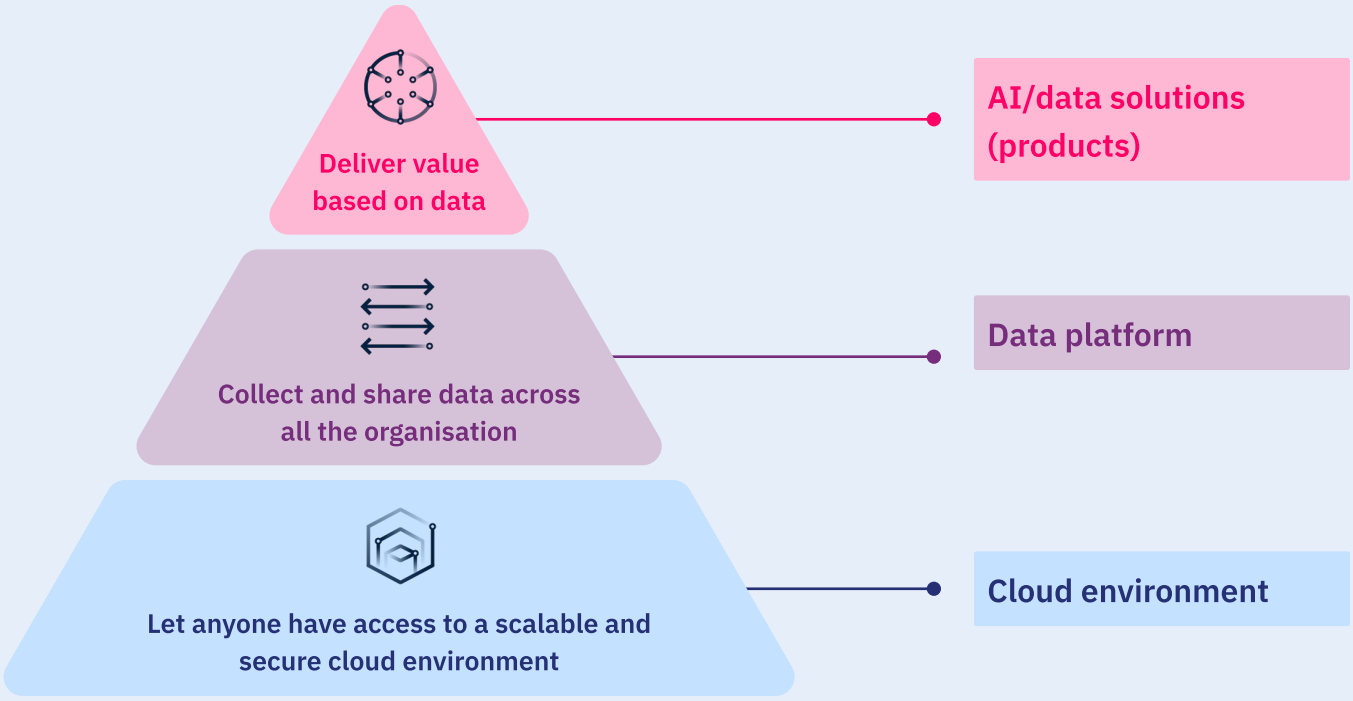


Data platforms have played a central role in the advancement of analytics tools, especially in the development of more sophisticated solutions based on large volumes of data. This advance is driven by the ability of these platforms to support complex techniques, such as machine learning, predictive modeling, and prescriptive analyses, which are the foundation of data-driven decision-making.

Unlike traditional Business Intelligence (BI) solutions, which focus on descriptive reports and retrospective dashboards, advanced analytics tools offer more precise and proactive insights. They are capable of performing data mining, generating predictions, and recommending actions, allowing for a higher level of business intelligence.

For these solutions to work efficiently, a modern, scalable, and cost-effective data architecture is essential. This architecture usually relies on three main levels:

- Cloud infrastructure
- Data Platforms
- Analytical solutions



AI/data solutions (products)

Assets to build: Develop methods and standards to build production ready AI products that guarantee the accuracy of the results over time and its resilience.

Actions to perform: Build ML ops pipeline to track the lifecycle of the AI model and build a CICD to build a product aligned with production requirements.

Data platform

Assets to build: Build a single source of truth that aggregates all the data sources. Expose and document them to make it accessible to everyone.

Actions to perform: Collect, clean, aggregate and expose data across your company. Leverage on your data governance strategy to share the right data to the right person.

Cloud environment

Assets to build: Manage your Organisation in order to setup security policies. Interconnect your on premise network with your cloud provider.

Actions to perform: Design an infrastructure. Build it to include core security, network and monitoring features.

At the first level, the cloud provides the necessary foundation for scalability, elasticity, and security. Governance policies, access control, encryption, and continuous monitoring are fundamental to ensuring that all users have access to a secure, reliable environment aligned with compliance and performance requirements.

The second level refers to the centralization of data in Data Platforms. The platforms are responsible for orchestrating the collection, cleaning, transformation, and availability of data throughout the organization. They function as the heart of the architecture, ensuring that data is accessible, standardized, and ready to feed the analytical models, always in compliance with governance guidelines.

At the final level, with the data already processed and structured, the analytical algorithms and AI models come into play. This is where they are truly transformed into value, with machine learning models that are trained, tested, and put into production to generate reliable predictions, detect hidden patterns, in addition to supporting strategic decision-making in real time.

This journey, from storage to predictive analysis, is only possible when the data architecture is well-defined, aligned with the company's strategy, and supported by a robust and modern platform.

It is this foundation that makes the advanced analytics and artificial intelligence success cases that we are starting to see more and more frequently in the market viable.

GenAI and Analytics: Success Stories in the Market

Generative artificial intelligence (GenAI) is rapidly transforming the way companies make decisions, create products, and relate to customers. As a result, its prominence is revealed when combined with robust analytics and a consolidated data base. In this chapter, we will explore how organizations in different sectors are integrating GenAI and Analytics into their operations and what results they are achieving with this synergy.

Historically, most companies in Latin America structured their data platforms with a focus on operational reports, dashboards, and descriptive analyses, with some others moving towards the use of Machine Learning. These traditional architectures are usually based on data warehouses, pipelines, and automation processes. Their main objective is the centralization, improvement in quality, and availability of data.

With the arrival of GenAI, however, new demands for Data Platforms emerge. Real-time ingestion, automated governance, semantic catalogs, and integration with LLMs allow the distances between the technical world and business users to be increasingly shortened. By translating complex contexts, interpreting intentions in natural language, and even generating SQL queries automatically for real-time responses, these new capabilities reposition the data platform as a direct facilitator of decision-making, expanding its strategic role in organizations.

Categories of Generative AI Use Cases:

Interact

*Use LLMs to **understand and analyze** data and requests in natural language*

Productivity Agent

Customer & Sales bot

Analytics Agent

Create

*Use LLMs to **create outputs ex nihilo**, whatever their format and medium*

Media content creation

Hyper Personalized targeting and campaigns

Financial Reports Generation

Automate

*Use LLMs to augment people by **optimizing and automating** tasks*

Automated Data processing

Information Extraction

Data Quality Enrichment

The use cases of GenAI can be organized into three major categories, according to their purpose and application:

Interaction | **Creation** | **Automation**

In all these categories, the use of Large Language Models (LLMs) is fundamental. These models act as intelligent intermediaries, interpreting commands and data in natural language to perform analyses, generate content, or execute tasks with precision.

The **Interaction** category includes agents that communicate with users or with each other, playing roles in areas such as customer service, sales, and productivity. A practical example is the development of sales chatbots for BNP Paribas and productivity agents implemented at Orange, both created by Artefact.

In the **Creation** category, systems are capable of generating content autonomously, from texts and images to more complex reports. These solutions were successfully applied to clients such as **Red Bull**, with media content generation, and at **Orange**, with financial report automation.

Finally, the **Automation** category focuses on the execution of repetitive or operational tasks without human intervention. This includes everything from intelligent data processing to quality testing. Artefact developed solutions in this line for companies like Legrand and, again, for **Orange**, showing the versatility of GenAI in critical operations.

Chatbot customer care

HelloBank

Challenges

Responding to customer requests consumes most of customer care teams' time

Current automatic response tools are unsatisfactory (imprecise, non-personalized responses)

Solution

Development of a Generative AI ChatBot designed to help customers solve problems autonomously

RAG on business knowledge databases

Fully deployed with open source capabilities

Results

Estimated €5m in saved costs over 3 years (optimization of customer care teams)

+15% use of connected spaces (mobile & web)

> **For BNP Paribas, a customer service chatbot was developed with the objective of automatically responding to different requests, offering effective solutions autonomously.**

The developed tool uses the Retrieval-Augmented Generation (RAG) technique, integrated into a knowledge base about the bank's business, which allowed for more precise and contextualized responses. As a result, the project generated an estimated saving of €5 million in three years, mainly through the optimization of customer service teams. Additionally, a 15% increase in the use of connected spaces is expected, reflecting an additional gain in operational efficiency.

Data Foundation: Global Healthcare Company

Context

Sanofi has an ambitious plan to reach the e-commerce edge by 2025... To achieve this objective we started building the big data foundation that supports the accelerated e-commerce growth and enables AI use cases

Objective

Centralized strategic data sources for e-commerce in a single source of truth enabling access to and cross-referencing of multiple databases in real time

Enablers

- ✓ **DC:** Understand business needs, design business cases, and connect business decisions with data
- ✓ **DE:** follow DevOps best practices, using lean, link data sources and make the data infrastructure globally scalable

> **Within the scope of Data Platform projects, at Sanofi, a pharmaceutical company, we developed a data foundation with a focus on big data, with the objective of centralizing all information related to e-commerce in a single reliable source.**

This platform enables data-driven decision-making, through a dashboard with more than 30 different views of strategic information. The initiative was essential to support the company's ambition in e-commerce, accelerate growth in this channel, and enable the use of artificial intelligence solutions, with the goal of achieving digital excellence by 2025.

Data Platform: Centralize and share data o enable IA use cases

Objective

Centralize and share data for Beauty Tech use cases. Provide services useful to Beauty Tech use cases in a use case agnostic approach. Accelerate implementation of new use cases

Design

Three core principles to drive platform build:

- **Operability**
- **Maintainability**
- **Evolutivity**

Achievements & Learnings

Build of an end-to-end ingestion and exposition pipeline released to production after 5 months

2 UCs sourced from the Data Platform

- **Roadmap prioritized per use case**
- **Scalable by design**
- **Data Governance at the beginning**
- **Ecosystem RACI defined**

> **For L'Oréal Paris, a Data Platform focused on information centralization was developed, with the objective of enabling the development of AI use cases aimed at Beauty Tech.**

The solution's architecture was built based on three fundamental principles: operability, maintainability, and evolvability, ensuring a solid and scalable foundation. As a result, an end-to-end data ingestion and exposure pipeline was delivered, put into production in just five months. Additionally, two use cases have already been developed and activated directly from the new platform.

Through real cases and achieved numbers, it is possible to observe how modern data platforms not only make this integration viable but also enhance its scale and impact. The objective is to highlight how the right combination of technology, data, and strategy can generate measurable innovation and tangible value for businesses.

The Convergence of Data and AI: Transforming Decision-Making

The union between data and AI is changing the way organizations make decisions. Until the last decade, data analysis was, for the most part, based on reports and historical indicators. Currently, with the advancement of AI algorithms, it is possible to predict scenarios, automate decisions, and act in real time based on precise insights.

By integrating structured and unstructured data with AI models, companies can extract value from previously underutilized sources, such as customer interactions, images, videos, and text.

This fusion allows for a global view of the business and consumer behavior, which drives strategies to be more assertive, personalized, and scalable.

An example of this transformation is Amazon, which has been using GenAI to improve everything from logistical planning to same-day package delivery.

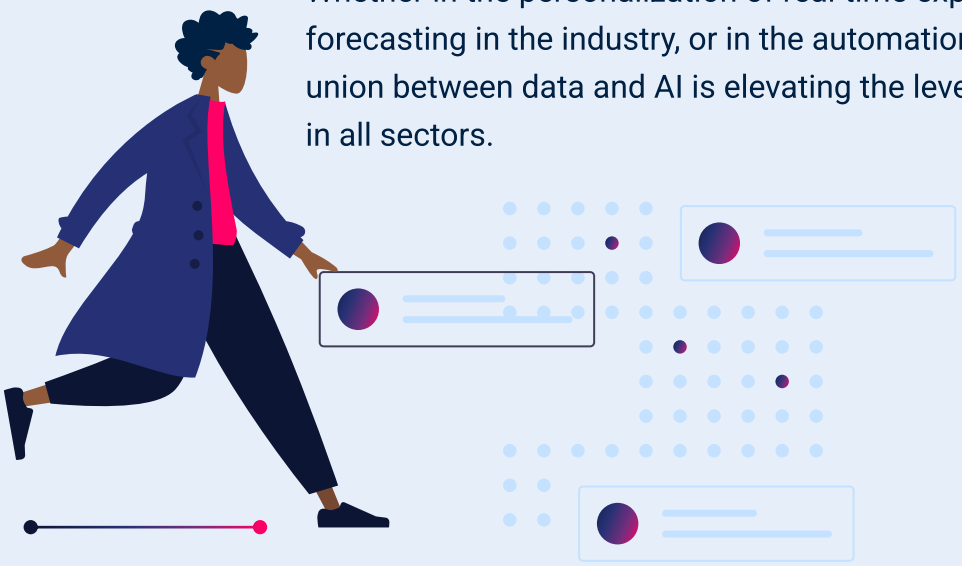
Based on large volumes of data, the company can forecast demand, optimize its supply chain, and define more efficient routes. Additionally, it uses AI to strategically decide where to store products, also to reduce its carbon footprint through a smarter logistical network. In distribution centers, autonomous robots equipped with AI help in the movement of items, reducing the physical effort of employees and increasing the agility of operations. This example demonstrates how the convergence between data and AI goes beyond accelerating decision-making, as it also transforms the daily operational life of organizations.

Therefore, AI models powered by robust data platforms act as a catalyst for digital transformation, as they identify patterns, detect anomalies, recommend actions, and, in many cases, execute tasks with autonomy. This new approach accelerates decision-making, reduces errors, and strengthens the competitiveness of companies.

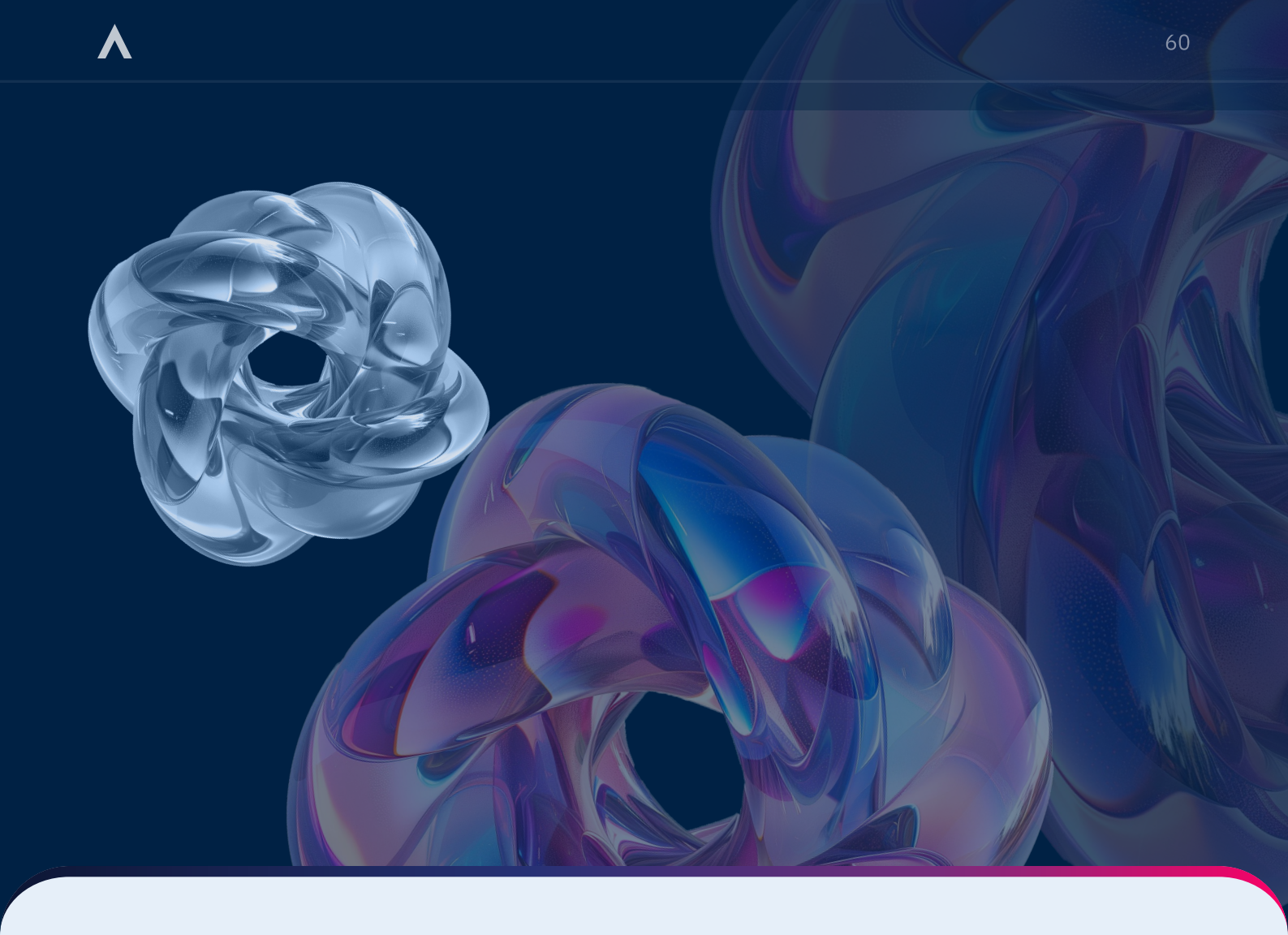
This impact can happen at different organizational levels, from operational to strategic, through the automation of repetitive tasks, support in planning and optimizations, helping managers to identify bottlenecks, adjust resources, and even align operations with strategic objectives with more precision.

Real cases demonstrate how this convergence is already generating an impact.

Whether in the personalization of real-time experiences in retail, in demand forecasting in the industry, or in the automation of financial decisions, the union between data and AI is elevating the level of operational intelligence in all sectors.



Therefore, the era of isolated intuition is being replaced by decisions based on evidence, context, and predictive capacity. The convergence between data and AI is not just a technological trend, but also a new paradigm for business management and innovation.



CHAPTER 4

Challenges and Solutions in Implementing Data Platforms

Main Challenges in Modernizing the Data Environment

> **The modernization of the data environment faces a series of challenges that require strategic planning and multidisciplinary involvement. One of the main obstacles is the complexity of the legacy infrastructure, which is often not prepared to support modern cloud-based technologies or data-oriented architectures.**

Additionally, there is a significant difficulty in the integration of dispersed data, often spread across multiple areas within the company and in different formats. Another critical factor is the lack of data governance and quality, which compromises the reliability and analytical value of the information. The scarcity of qualified professionals in modern technologies, such as data engineering and cloud architecture, means many companies still do not have the necessary know-how to assemble a team capable of leading this change. Finally, there is a natural organizational resistance to change, which requires cultural change management to align teams, processes, and objectives around a data-driven vision.

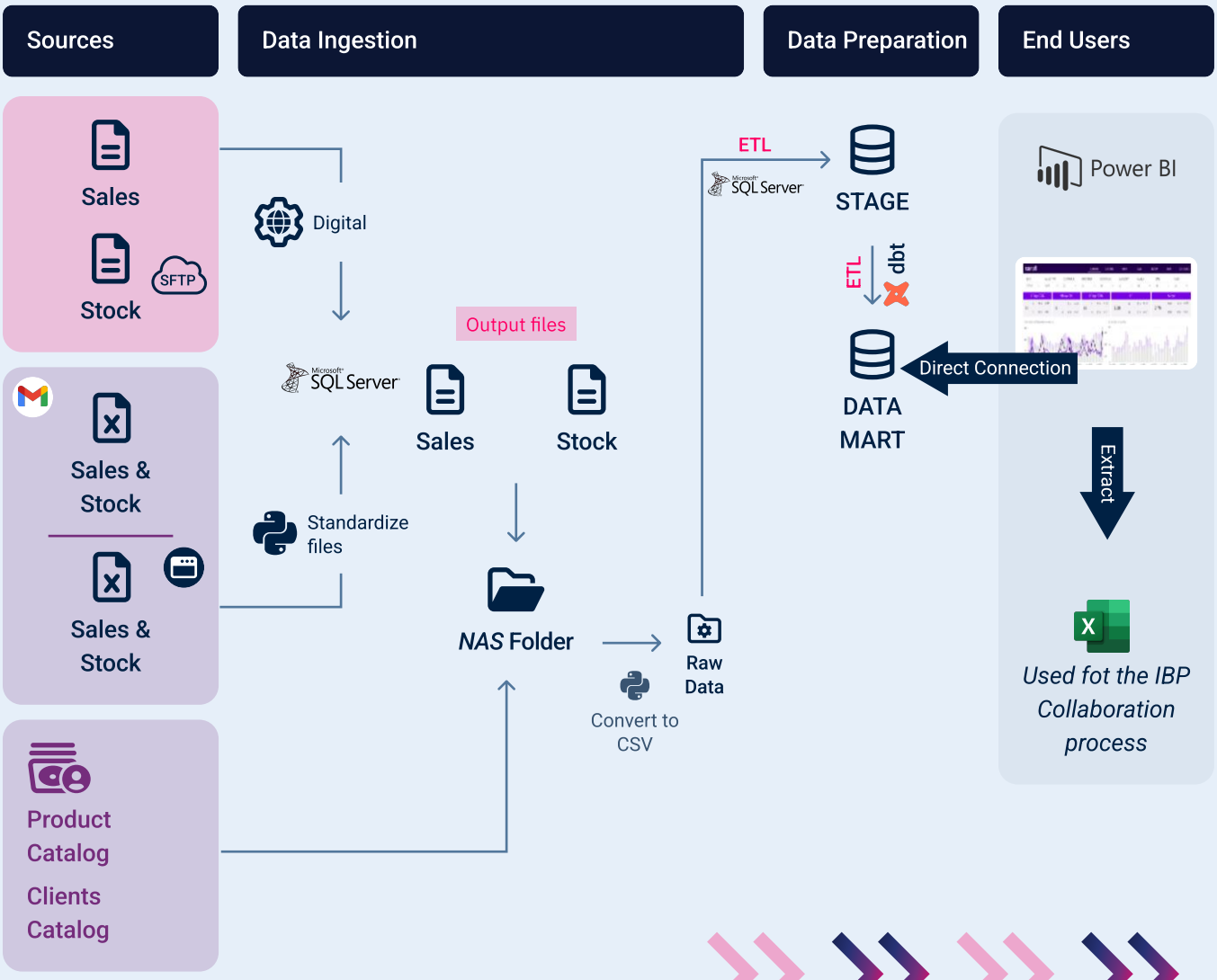
Strategic Solutions to Overcome Common Barriers

To overcome the complexity of the legacy infrastructure, an effective strategy is to adopt a hybrid approach, which combines existing systems with modern cloud-based solutions, allowing for a progressive and secure migration. This strategy must go hand in hand with data integration, through the use of modern tools (such as Databricks, AWS Glue, or Cloud Dataproc), which make it possible to extract, transform, and load data efficiently and, whenever possible, with the best governance.

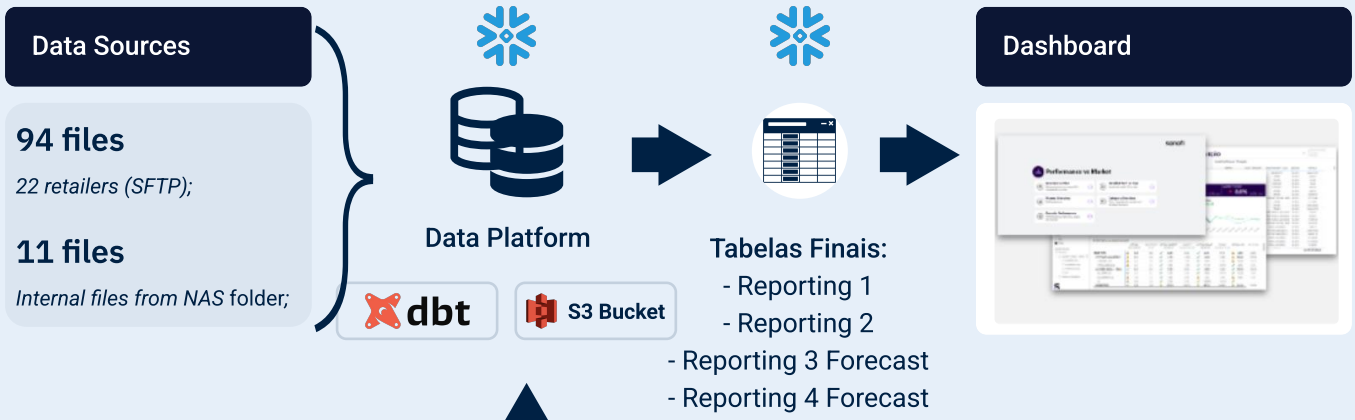
We can see an example below of a migration that started from a decentralized process, with a legacy structure and a certain degree of governance, transforming into a centralized process, with high availability, easier maintenance, version control, and future expansion capacity.

The main inputs for this change are good governance, a trained technical team, and organizational engagement to support the adoption of a data-driven culture, ensuring the sustainability of the transformation in the long term.

LEGACY STRUCTURE



NEW DATA PLATFORM



Value from this project:

Tech value

- ✓ Data platform infrastructure
- ✓ Data quality trackers
- ✓ Technical documentation & training session

Business Value

- ✓ Automation of ingestion and transformation process
- ✓ Impulse the data democratization ambition

Data governance is a key factor in dealing with the lack of data quality, making it necessary to implement structured management policies, supported by data catalogs and continuous monitoring and improvement processes. The scarcity of qualified professionals can be faced with internal training programs and strategic partnerships with specialized consultancies.

Finally, to reduce resistance to cultural change, it is essential to apply structured change management practices, engage leaders, promote a data-driven culture, and demonstrate practical gains from the first stages of modernization.

The Role of Data Governance and Security in Data Architecture Success

Data governance is one of the fundamental pillars for the success of any modern data architecture.

It ensures that data is managed with quality, consistency, and traceability throughout its entire life cycle. This includes everything from defining clear standards and policies to implementing processes for classification, cataloging, and access control—an element that is directly intertwined with information security.

With well-structured governance, it is possible to increase confidence in the information used by business areas, accelerate the adoption of analytics and AI solutions, in addition to reducing rework and operational risks.

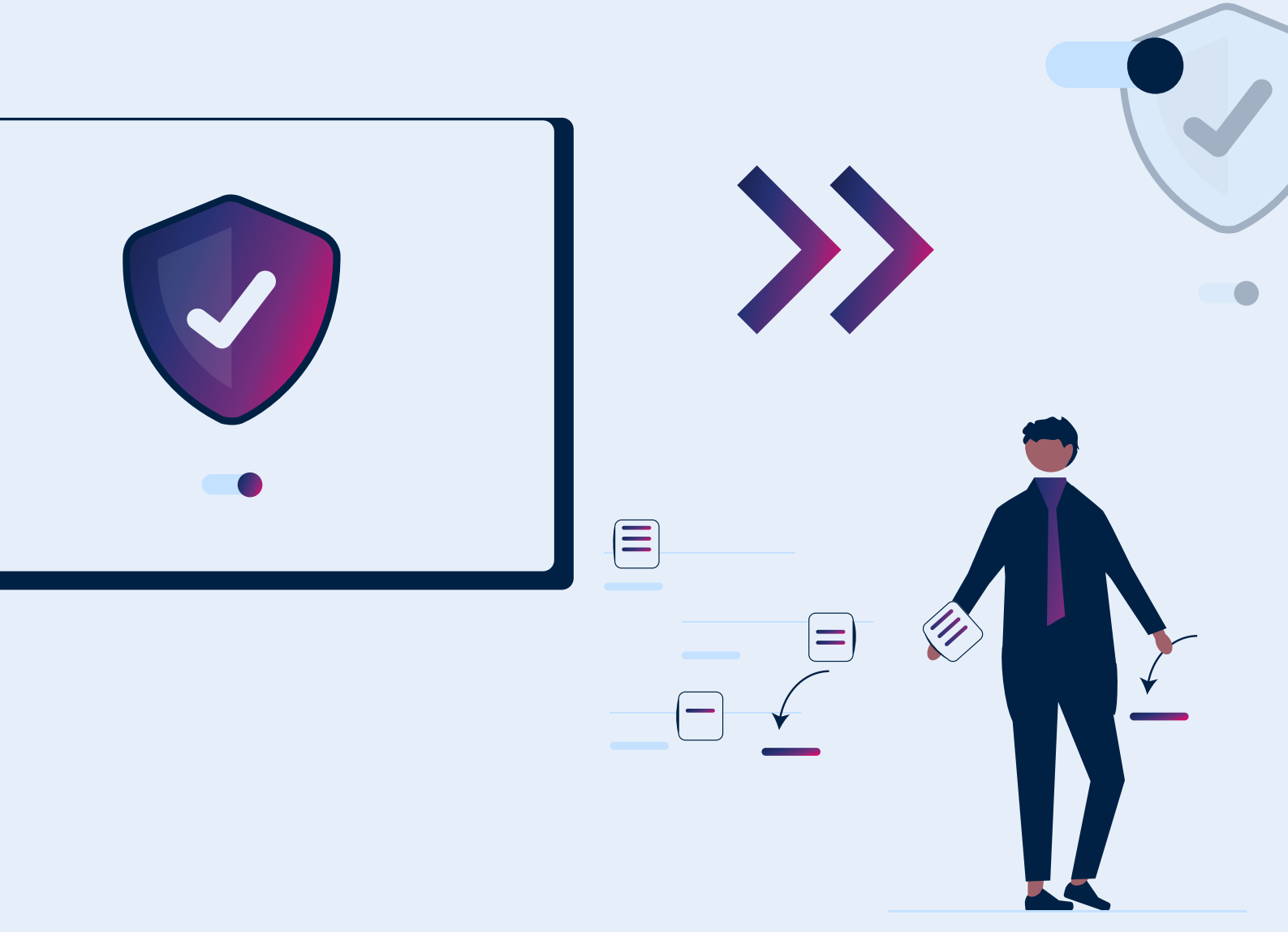
In this context, information security acts as an essential complement, protecting the data ecosystem against undue access, leaks, and unauthorized manipulation.

With the increase in the sensitivity and volume of data processed—especially in distributed and cloud-based environments—the adoption of a **Zero Trust Architecture (ZTA)** approach becomes essential. The Zero Trust model operates under the principle of "never trust, always verify," requiring continuous authentication and granular access policies based on roles and contexts such as location and device, thus aligning with the established governance guidelines.

Sources: DataGuard | Microsoft Learn

Furthermore, ZTA is not limited to traditional perimeters but protects sensitive data even in hybrid infrastructures, mitigating risks such as compliance violations. Compliance with regulations such as **LGPD** and **GDPR**, for example, depends on both robust governance practices, for traceability and standardization, and security, for minimum privilege control and active protection.

*Thus, by combining data governance with Zero Trust security, organizations not only strengthen their architecture but also create a **reliable environment** for innovation and sustainable growth—where data is accurate, accessible only to those responsible, and aligned with the most demanding regulatory standards.*



CHAPTER 5

Strategy for Implementing

The Strategy for Implementing Data Platforms: **Diagnosis and Strategic Planning**

> **The implementation of a Data Platform requires a strategic approach and careful planning. For the organization to succeed in this process, it is essential to perform a diagnosis of the current data scenario and then draw up a detailed execution plan.**

Below, we present a practical step-by-step guide to diagnose the company's data scenario and plan the implementation of a Data Platform.

STEP 1 DIAGNOSIS OF THE CURRENT DATA SCENARIO

Before starting any implementation, it is crucial to understand where the company is in relation to the maturity of its data and how this data is being used in its operations.

An in-depth diagnosis helps identify gaps and opportunities, in addition to ensuring that the implementation process is guided by the real needs of the business.

CONCRETE ACTIONS FOR THE DIAGNOSIS:

Mapping Data Sources

Identify all existing data sources in the organization (internal and external), such as databases, ERP, CRM systems, spreadsheets, and unstructured data (such as emails and documents). A complete survey of these sources allows for a clear view of where data is stored and how it is being used.

Data Governance Analysis

Evaluate data quality in terms of integrity, accuracy, security, and compliance. This includes checking for the existence of data governance policies, such as access rules, data cleaning processes, and auditing. Companies that do not have structured governance face challenges such as outdated data and decisions based on wrong information.

Decision Process Evaluation

Analyze how data is used in strategic, operational, and tactical decisions. Ask yourself: are decisions based on real-time data or on historical reports? Is there a lack of integration between departments that prevents a holistic view of the business? Identifying these points will help to understand the gaps in data usage.

Current Technical Capacity

Check the existing data infrastructure (legacy systems, data lakes, data warehouses, etc.) and evaluate whether it supports current and future needs. This includes analyzing the scalability of solutions, integration between platforms, and flexibility to adopt new technologies, such as AI and machine learning.

RESULTS OF THE DIAGNOSIS:

- **Mapping data gaps:** The diagnosis should make it clear where the failures in data management and use are (for example, lack of integration between systems, outdated data, or manual collection processes).
- **Critical points to be addressed:** The analysis must identify problems such as the lack of governance or the use of outdated technologies that hinder data scalability.

STEP 2:

DEFINING BUSINESS OBJECTIVES AND PRIORITIES

Once the diagnosis is clear, the next step is to align business objectives with the implementation of the Data Platform.

This alignment is fundamental to ensuring that the data architecture is not just a technical solution, but a facilitator for achieving the company's strategic goals.

CONCRETE ACTIONS FOR DEFINING OBJECTIVES:

Understanding Strategic Objectives

Conduct interviews with executives and stakeholders to understand what the company's main goals are, such as increasing revenue, reducing costs, improving the customer experience, or innovating new products. From there, define how data can be used to drive these objectives.

Prioritizing Use Cases

Identify the most urgent use cases, such as personalizing the customer experience, predictive analysis to optimize sales, or improving operational efficiency. Focusing on high-impact use cases, such as process automation or creating real-time insights, helps ensure that the investment brings results quickly.

Establishing KPIs

Define key performance indicators (KPIs) that will allow measuring the success of the Data Platform implementation. Examples of KPIs include an increase in the speed of decision-making, an improvement in the accuracy of reports, or the financial impact of operational improvements.

CONCRETE EXAMPLE:

A retail company can define the strategic objective of increasing revenue through the personalization of offers. In this case, the Data Platform can be planned to integrate customer, transaction, and purchasing behavior data in real time, allowing for more relevant promotions, which can result in a 15% increase in sales.

STEP 3

DEFINING THE NECESSARY DATA ARCHITECTURE

With the objectives and priorities defined, it is time to design the data architecture that will allow these results to be achieved.

The data architecture needs to be chosen based on the company's specific needs and the identified use cases.

CONCRETE ACTIONS FOR DEFINING THE ARCHITECTURE:

Choosing the Appropriate Platform

Decide whether the solution will be on-premise, in the cloud, or hybrid. Platforms like Amazon Redshift, Google BigQuery, or Microsoft Azure Synapse are popular options for companies seeking scalability and flexibility. The choice of platform should consider the volume of data, the complexity of the processes, and the need for integration with other solutions.

Architecture Design

The data architecture can include components such as Data Lakes, Data Warehouses, Data Marts, and data integration solutions (ETL). For example, an e-commerce company can implement a Data Lake to store raw user behavior data and a Data Warehouse to consolidate transactional and financial data, making them accessible for operational and strategic analyses.

Ensuring Governance and Security

Include data governance tools (such as defining access, auditing, and compliance policies) to ensure that data is always protected and in compliance with regulations (example: GDPR).

STEP 4

IMPLEMENTATION PLANNING

With the architecture defined, it is time to plan the project's execution. The planning must be structured to reduce risks and ensure that business goals are achieved within the deadline and budget.

CONCRETE ACTIONS FOR IMPLEMENTATION PLANNING:

Implementation Roadmap

Create a detailed roadmap that describes all phases of the implementation, from data preparation to integration with external systems. Establish clear milestones, such as the completion of a platform prototype, the integration of data sources, or the implementation of advanced functionalities such as predictive analysis.

Phased Execution (Pilots and Iterations)

Start with a pilot or a prototyping phase to test the platform's viability in a controlled environment. This allows for quick adjustments before a large-scale launch. The use of agile methodologies allows for continuous adjustments based on feedback.

Team Capacity Building and Engagement

Ensure that the teams involved are trained to use the new data platform. The project's success depends on the adherence of internal users, who need to understand the value of data and how to use it to generate insights.

STEP 5 MONITORING AND CONTINUOUS ADJUSTMENTS

After implementation, it is important to monitor the results and make adjustments as needed. Monitoring should focus on both the platform's performance and the defined KPIs to ensure that business objectives are being achieved.

CONCRETE ACTIONS FOR MONITORING AND ADJUSTMENTS:

Performance Monitoring

Use monitoring tools to evaluate platform usage, system performance, and data integrity. This includes monitoring query latency, data storage efficiency, and the availability of data sources.

Business Impact Evaluation

Periodically review the KPIs to check if the platform is generating the expected results, such as increased sales, reduced costs, or improvements in the customer experience.

Iteration and Continuous Improvements

Based on the monitoring results, implement continuous adjustments to improve the platform, add new use cases, and expand data integration.

CHAPTER 6

Impact Examples

Examples of Successful Transformations with Data Platforms and Agentic Platforms AI

This chapter explores how leading companies in their sectors are driving real transformations through well-structured data platforms and efficient governance. Based on projects implemented by Artefact, concrete examples are presented that show the tangible results obtained in different industries, from beauty and retail to telecommunications, pharmaceuticals, and financial services. It closes with the practical lessons learned from these implementations, which are four fundamental pillars for the success of a sustainable data strategy.

Transforming Companies with Data: Real-World Success Stories

The journey to becoming a data-driven organization involves much more than technology: ***it requires vision, structure, governance, and cultural change.***

Several global companies have successfully traveled this path with the support of Artefact, achieving expressive gains in efficiency, margin, and autonomy in decision-making. Below, we highlight some real cases that illustrate how the implementation of data platforms and the activation of AI use cases can transform businesses in different sectors.

L'ORÉAL | **N°1 Beauty Tech Partner**

- ✓ 4 years continuous support (20 Artefact members)
- ✓ L'Oréal group data platform design & implementation
- ✓ Lead Data governance program: 3 years, 18 data domains & 30 countries
- ✓ Data use cases deployed at scale

Ex.: *Trendspotter, 1,000+ users, 3 business lines, 7 languages*

Hundreds M€ of use cases enabled

Carrefour – Google Lab

- ✓ Management of the lab for 2.5 years with 12 Artefact data experts
- ✓ 6 use cases implemented at scale: eCommerce orders forecasting, PoS assortment customisation, promotions volume forecasting, etc.
- ✓ 40 BI solutions delivered to business teams
- ✓ 300 Carrefour employees trained

300 Employees trained
100 M€ incr. Margin

AI Acceleration Program

- ✓ 6+ years of continuous & growing support on data services
- ✓ 45 Artefact people since June 2021
- ✓ 20 use cases under development: product recommendation engine, augmented field technician etc.

Meta: 120 M€ de lucro líquido
+3 pontos NPS

HR Data Platform

- ✓ Build HR data platform containing highly sensitive information at the LVMH group scale
- ✓ Native alignment with security requirements in order to ensure the safety of this data on the cloud

Ongoing project

Advanced Analytics Transformation Program

- ✓ Build and run Fortenova Data Lab: data platform, data governance, use cases, change management
- ✓ Design and implementation of analytics platform on GCP and integration of Data Governance by design
- ✓ Develop and deploy AI/Analytics use cases in Retail: Assortment optimization, Demand Forecasting, Promotion optimization, BI solutions

Ongoing project

L'ORÉAL

Data as a Global Strategic Pillar

Over four years of continuous partnership with more than 20 dedicated experts, we supported the design and implementation of the L'Oréal Group's global data platform, in addition to leading the data governance program in 30 countries and 18 domains.

Among the implemented use cases, the Trendspotter stands out, a market intelligence solution with more than 1,000 users in 3 business lines and 7 languages, demonstrating the scalability of the data-driven approach.

This journey consolidated data as strategic assets, integrating governance and technology to support continuous innovation.

CARREFOUR

Data at the Center of Retail Decisions

With 2.5 years of data lab management and a dedicated team of 12 Artefact experts, Carrefour was able to scale 6 high-impact use cases, including e-commerce order forecasting, point-of-sale assortment personalization, and promotional volume forecasting.

Additionally, 40 BI solutions were delivered to different business areas, and more than 300 employees were trained, expanding the analytical culture throughout the organization. The result was a more agile, assertive, and insight-driven retail.

ORANGE

Data and AI at Scale with a Long-Term Strategy

With more than 6 years of continuous support and 45 experts involved since 2021, Orange has been accelerating its data and AI transformation program. Currently, 20 use cases are under development, including: product recommendation engines and augmented AI field solutions.

The strategy also includes robust governance initiatives, a scalable infrastructure, and integration with business teams, consolidating Orange as a reference in data-based digital transformation.

Strategic Data Migration to the AWS Cloud

Context

Part of **ANBIMA product portfolio** consists of dataset and analysis of different investment funds. Due to change in legislation, the way these datasets are stored and organized will need to change **by the end of march/24.**

Objective

Most of the legacy is built on complex GCP datasets and since most of these will need to be adapted into a new format, **ANBIMA wants to migrate them to a more simple and easy to use AWS environment.** This environment also needs to be built.

Challenges

In order to migrate from GCP to AWS all new table schemas need to be built from scratch

An easy to use architecture needs to be built using AWS and implemented alongside Databricks

ANBIMA uses multiple tech partners and coordinating with them is a must

> **In the financial sector, we helped ANBIMA prepare for regulatory changes, migrating its investment data sets from a complex GCP environment to a new simplified architecture on AWS with Databricks.**

The project required the creation of schemas from scratch, the definition of an easy-to-use architecture, and coordination with multiple technology partners. The delivery enabled not only regulatory compliance but also gains in performance and autonomy in data management.

These cases show that, when well-planned and executed, a data strategy can generate hundreds of millions in business value, whether through gains in margin, operational efficiency, loss reduction, or faster and more accurate decisions. More than technological projects, they are stories of real transformation, driven by data, enabled by people, and sustained by processes.

Visible Impacts on Decision-Making and Operational Efficiency

The implementation of data platforms has caused a transformation in the way organizations make decisions and operate on a daily basis. By centralizing, integrating, and making information available in real time, these platforms break with old paradigms based on manual processes, intuition-based decisions, and analyses limited by outdated or ungoverned data. Consequently, a new dynamic for decision-making emerges, being more agile, precise, and evidence-driven, accompanied by substantial gains in efficiency in various areas of the company.

This infrastructure allows immediate access to interactive dashboards and dynamic reports that reflect operational reality with precision. This means that strategic decisions can be based on concrete and up-to-date data. Similarly, tactical decisions are now made based on information that reflects the current behavior of the market and consumers.

In addition, operational efficiency and support for decision-making are benefited by integrating and automating workflows. With data platforms, it is possible to eliminate redundancies, minimize manual errors, and reduce the time needed to perform critical tasks.

*A good example is the project implemented at **Danone**, in which the consolidation of a unified P&L view of three business units was carried out.*

Fed by data from different source systems, it is able to calculate the rules, previously aligned in the mapping phase, and make the information available in a dashboard composed of different KPIs for the Marketing and Finance sectors.

Build a cross-BU unified vision of the P&L

Food and Beverage Company

Challenge

How to get from fragmented excel files to a unified & reliable vision of the P&L for the 3 Business Units?

Achievements

Reduction in manual work time for P&L construction.

Orchestration of the P&L construction workflow with the impacted teams.

Solution

Align the BUs on a common definition of P&L (lines, granularity…) and needs for analysis

Mapped the business & calculation rules for each line at SKU/Client level

Mapped the relevant data sources and design templates for manual inputs

Built data flows in Informatica Cloud, stored the data in Snowflake and expose the results in a PowerBI dashboard

Work hand-to-hand with the business and manage the daily tasks of +20 stakeholders (business, developers, IT…)

*With the migration to the GCP platform and the adoption of **Dataiku** as a self-service data preparation tool, the **predictability of Orange's analytical processes took a significant leap.***

Business teams that previously relied heavily on macros in Excel now have a robust and scalable environment to build, save, and reuse data flows, in a completely no-code way. The greater autonomy in data preparation allows teams to anticipate trends, explore data stored directly in BigQuery, and act on the basis of reliable and up-to-date insights. The result is a more agile operation, with better-informed decisions and a more efficient use of resources in various areas of the company.

The model has been transforming Orange's analytical culture, making data accessible, useful, and operational for everyone.

Tooling for BI/ AI & No code tooling

Data manipulation no code for business users

Context

Orange migration on GCP cloud platform

Ambition to democratize data access & transformations to make business teams more autonomous

Challenge

Offer a no-code data preparation self-service tool for non-technical users

Allow business users to build & save data workflows for periodic reuse (previously realized via Excel VBA macros)

Allow business team to use data stored in BigQuery directly

Approach

Requirements framing

- +20 users interviewed in many directions *(Finance, Network, Support, Marketing…)*
- Legacy use case analysis (+30)

Tools Benchmark

- 4 tools pre-selection: Dataiku, Alteryx, Google DataPrep × Trifacta, MS Power Query
- Benchmark on 5 axis: features, training course, license fees, integration on IT env. & processes

Demo & User Testings

- Demonstration of a selected use case on each tool benchmarked
- Hands-on sessions with +20 users

Recommendations, deployment & onboarding

- **Tool selection:** Dataiku
- Tool integration on Orange processes & IT env.
- Training courses launched

Results

- ✓ Dataiku for data prep. deployed in Orange IT ecosystem @ scale
- ✓ Access management processes & tools defined & deployed
- ✓ +50 daily users onboarded (new onboarding planned)

These examples make it clear that the adoption of an effective data platform is not just a technological matter, but a structural change in how the organization operates and competes. So that, they start to adjust their strategies based on real data and add more value to the results. The ability to act with intelligence, speed, and precision, in a dynamic scenario, becomes a decisive strategic differentiator.

Lessons from Practice: How to Move Forward with a Sustainable Data Platform Strategy

> **The adoption of a Data Platform, when well-conducted, translates into significant gains in efficiency, agility, and competitiveness. However, without a clear evolution structure, the project can become burdensome, underutilized, and misaligned with the company's objectives.**

Based on the accumulated experience in multiple sectors and observations of projects, we identified four fundamental pillars that support the advancement of any modern data platform: T**echnology, Ways of Working, Documentation, and Data Governance**. These pillars represent a practical guide to maximizing value and avoiding recurring errors.

1. **TECHNOLOGY**
   The Right Tools for the Right Development
2. **WAYS OF WORKING**
   Practical Collaboration between IT and Business
3. **DOCUMENTATION**
   Systematic and Standardized Processes
4. **DATA GOVERNANCE**
   Centralized Governance with Vision and Ownership

[13] Fonte: Conhecimento interno - Artefact

1

TECHNOLOGY
The Right Tools for the Right Development

Successful organizations in their implementations understood that technology is a means, not the end. One of the main lessons is that offering the technical team adequate and interoperable tools with the data ecosystem is essential to accelerate development.

The choice of the technology stack must be made with pragmatism: cloud solutions, modular architectures, and scalable platforms (such as the use of Dataiku on GCP) provide agility and allow dealing with different data formats and volumes efficiently.

The focus should not be on replicating market trends, but on adopting solutions that fit the organization's reality and offer a real gain in productivity.

Error to avoid

Investing in complex or inflexible technologies without evaluating internal maturity, integration with legacies, or total cost of ownership. This generates delays and waste.

2

WAYS OF WORKING
Practical Collaboration between IT and Business

Another evident differential was the improvement in the ways of working between technical and business teams. Mature organizations broke down silos and created collaboration models based on autonomy and mutual trust. By enabling the use of no-code/low-code tools, they empowered analysts and managers to explore data directly, creating reusable flows and accelerating value generation.

These teams also adopted agile approaches, with short delivery and validation cycles with the end user. This constant integration with the business increased the adherence of the developed solutions and drastically improved the adoption of the platform.

Error to avoid

Maintaining a centralized and reactive delivery model, where the business depends entirely on IT to extract value from the data. This creates bottlenecks and reduces the platform's strategic impact.

3

DOCUMENTATION
Systematic and Standardized Processes

Building a scalable platform requires consistent documentation and clear processes. Successful projects treated documentation as an essential part of development, with systematic versioning, traceability, and standardization practices.

By documenting flows, integrations, and usage policies, these organizations **reduced rework, facilitated the entry of new users,** and ensured **quality over time.**

The absence of documentation, on the other hand, made the platform's evolution difficult and limited its reuse and reliability.

Error to avoid

Neglecting the documentation of data assets and analytical processes. This compromises governance, creates dependency on key people, and hinders scale.

4

DATA GOVERNANCE

Centralized Governance with Vision and Ownership

Finally, data governance proves to be the most critical pillar for the platform's sustainability. Successful cases established a clear governance model from the beginning, with defined roles, quality metrics, access control, and security policies aligned with regulations such as LGPD and GDPR.

This centralized governance does not mean absolute control, but rather clarity about who is responsible for what, with transparent processes shared between areas. Without this, data quickly becomes duplicated, inconsistent, and unreliable.

Error to avoid

treating governance as a future or optional step. Platforms without governance become disorganized repositories, making decision-making based on reliable data unfeasible.

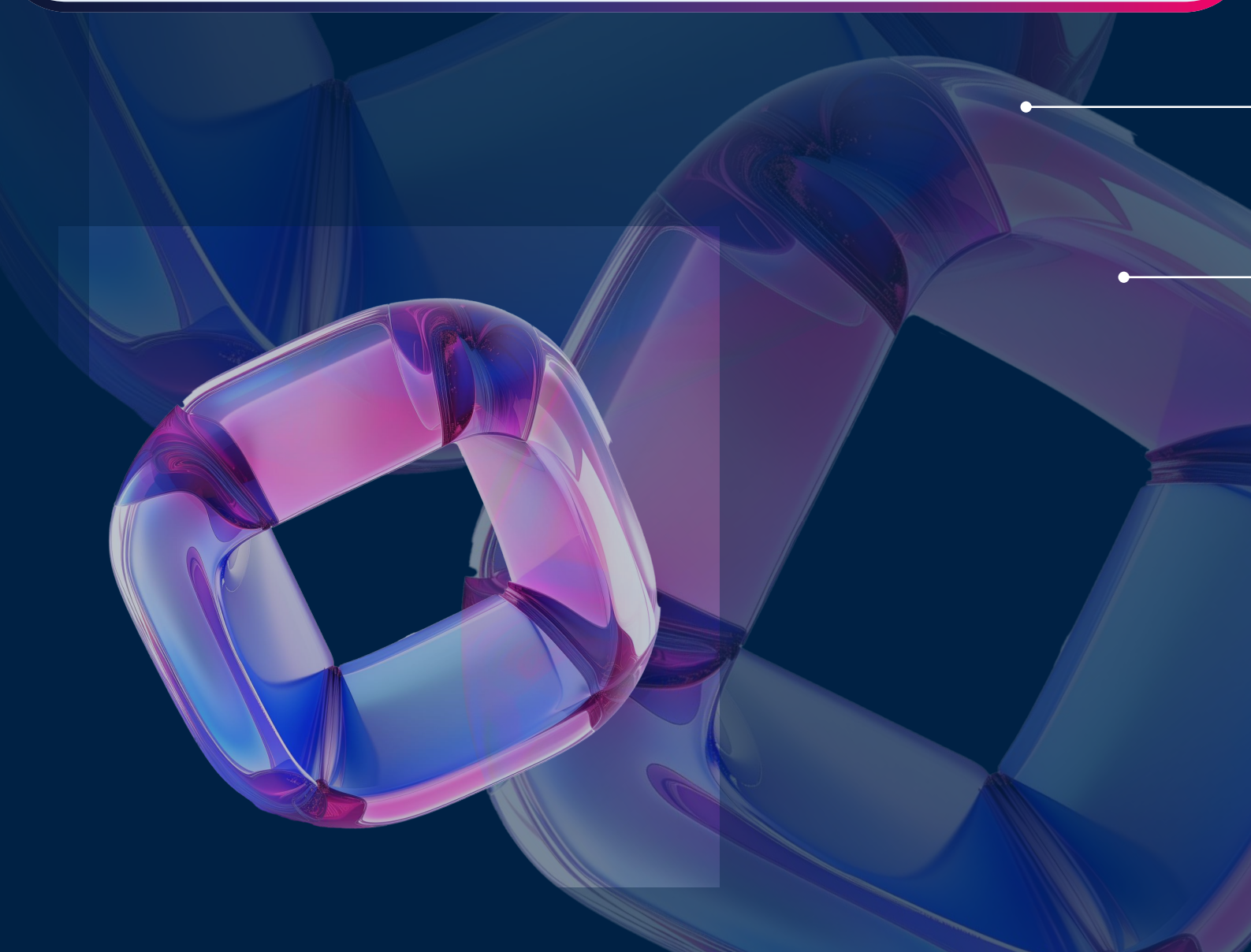
The implementation of a successful data platform requires more than a good architecture.

It requires strategic focus, alignment between teams, standardized processes, and a solid governance. The four pillars form a practical structure to guide decisions, prioritize resources, and maximize return on investment.

Learning from the successes and errors of other organizations allows for avoiding common traps and shortening the path towards analytical maturity. When well-planned and executed, a data platform is transformed into a true organizational asset.

CHAPTER 7

The Future of Data Platforms



The Future of Data Platforms - Agentic Platforms

> **Data architectures are no longer mere analytical repositories and are beginning to serve as real-time nervous systems for generative AI (GenAI) agents.**

In this section, we discuss how this change repositions ingestion, processing, and governance requirements—and why "agentic platforms" will be the new standard for data-driven organizations.

FROM ANALYSIS TO ACTING IN MILLISECONDS

1. Immediate response to events

Transactional chatbots, customer service copilots, or supply chain agents cannot wait for nightly batches: they need fresh data (sub-second) to decide, plan, and execute.

Emerging pattern: continuous ingestion via change-data-capture (CDC), IoT/Kafka streams, and WebSocket APIs, with "hot storage" (Redis, Pinot, Materialize) for sub-second queries.

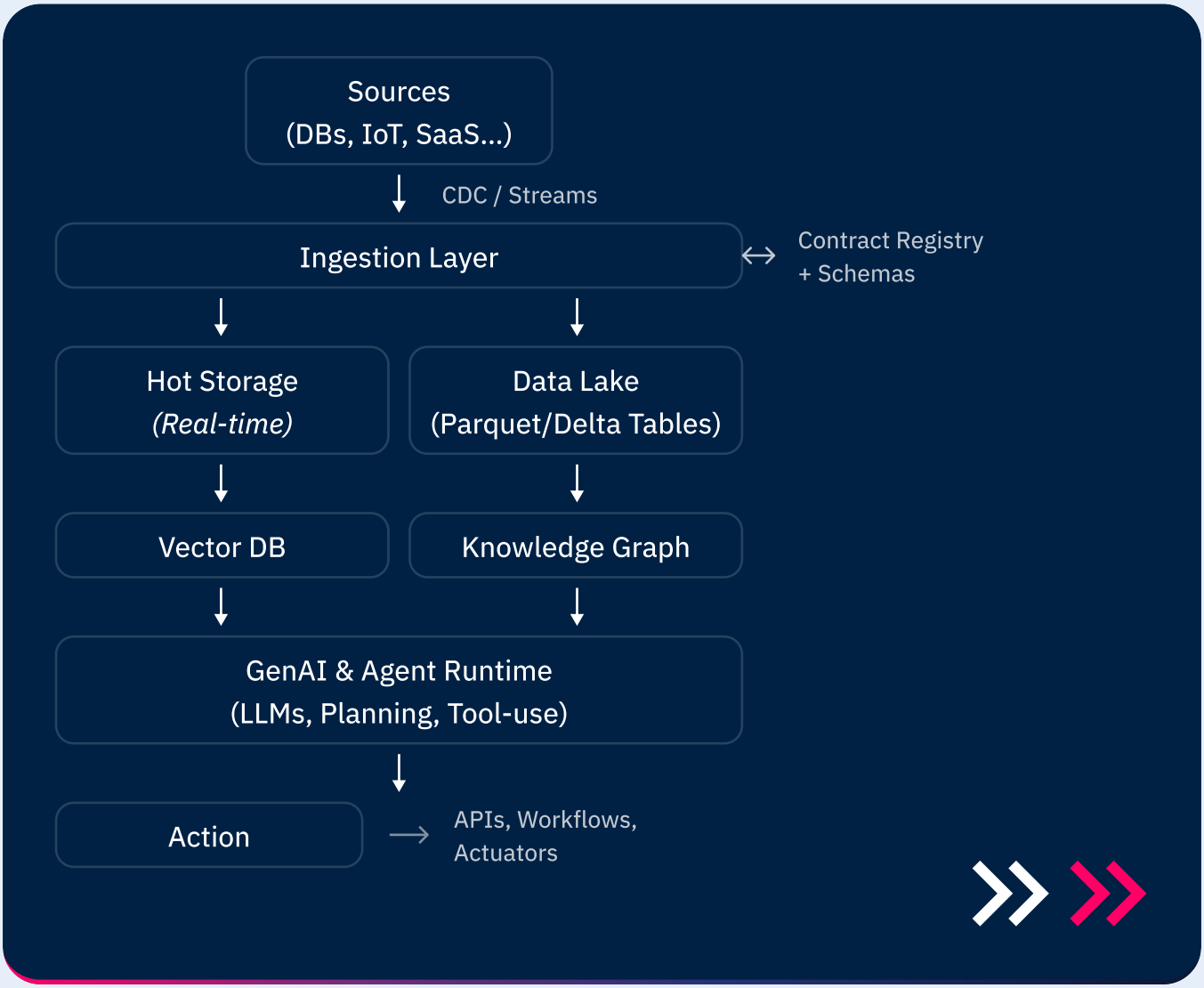
2. From human consumption to agent consumption

Dashboards and reports remain valid, but the first consumer is now a machine—an autonomous agent that orchestrates workflows, triggers transactions, or generates content.

Consequence: semantic layers need to be machine-readable (JSON-LD, RDF, contract-first APIs) and enrich the data with quality, trust, and temporal context metadata.

	ERA OF BI	AGENTIC ERA
Variety	Tabelas estruturadas	Texto, imagem, áudio, vídeo, telemetria
Speed	Lotes diários / hora	Streaming em tempo real
Useful Volume	GB → TB	TB → PB (para RAG e fine-tuning)
Semantics	Modelos dimensionais	Grafos de conhecimento + embeddings
Main Use	Descrição & diagnóstico	Planejamento & ação autônoma

- **Unstructured data becomes a first-class citizen:** clickstream logs, call center transcripts, inspection images, etc.
- **Hybrid storage:** Parquet Data Lakes (+Iceberg/Delta) coexist with vector stores (Milvus, Chroma, Weaviate) for Retrieval-Augmented Generation (RAG).



- **A double hot/cold path** guarantees low latency without sacrificing complete history.
- **Vector DB + Knowledge Graph** unite semantic similarity with explicit relationships—essential for agents to justify decisions.
- **Agent Runtime** contains orchestration, tool-usage, monitoring, and limits (governance).

1 Real-time Policies

- ✓ Data contracts versioned and validated on-stream.
- ✓ Drift alerts in embeddings or quality regressions of LLMs.

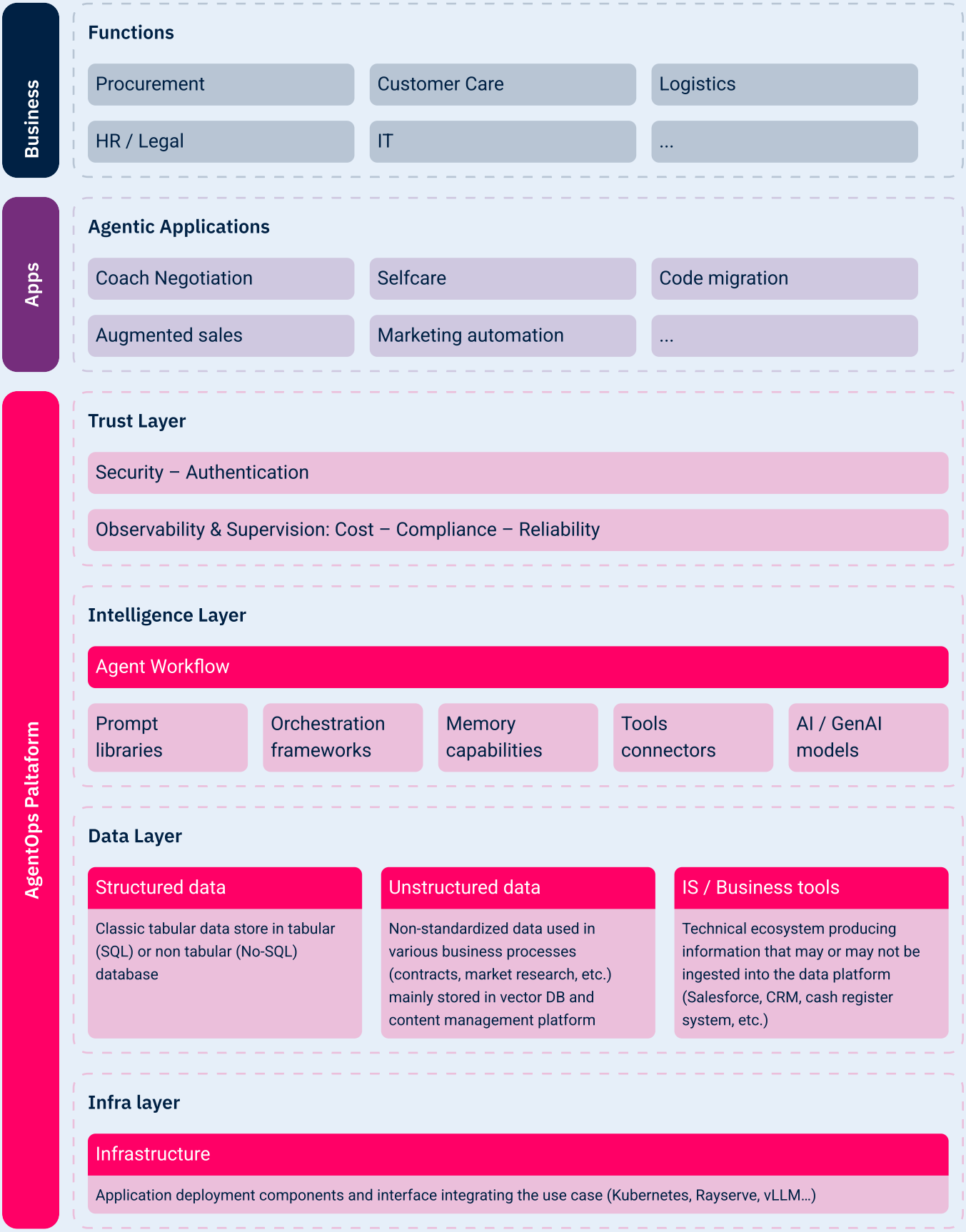
2 Auditing AI Decisions

- ✓ Complete traces: prompt → context (RAG) → response → action.
- ✓ Immutable hash in a ledger (blockchain or write-once storage) for compliance.

3 GenAI FinOps

- ✓ Inference costs tracked by agent, call, and business team.
- ✓ **Optimization:** caching, quantization, and dynamic routing between models.

AgentOps Platform relies on a modular stack combining infrastructure, data, intelligence, and trust layers



The adoption of agentic platforms brings a new layer of complexity because it overlays risks typical of streaming, multimodal data, and generative models on the already known risks of traditional BI. Next, I will describe, in a continuous text, the main risk groups, their symptoms, potential impacts, and practical countermeasures.

Data and LLM Quality and Accuracy

The first risk is critical hallucination: *when the agent generates a response without a reliable source or with a low confidence score (for example, probability < 0.30). This occurs, above all, if the context used in the RAG is outdated or if the model was "led" by an ambiguous prompt.*

The most visible symptom is usually negative feedback from users or incorrect actions at the endpoint (for example, wrong price or poorly directed email). Mitigation begins with mandatory RAG, symbolic or rule-based verification of responses, and prompt redundancy (self-check). The operational goal is to have at least 95% of responses with a traceable source and less than 1% of false positives.

Another point is data drift or concept drift. *When the distribution of events changes, the RAG recall drops and the agent starts to ignore new facts.*

Warning signs are the drop in recall@10 or significant variations in the cosine distance of the embeddings. It is essential to monitor this delta and incrementally re-index the vector store, in addition to retraining or applying LoRA at each cycle (for example, a 30-day moving window). A good practice is to keep recall@10 above 90% and a distance variation of less than 0.08.

Costs and Scalability

Agentic platforms can suffer from runaway costs—GPU or processed token spikes that blow the operational budget. This usually appears in hourly cost graphs or in cloud provider budget alerts.

To contain the problem, cache prompts, compress them, route calls to cheaper models when the request complexity is low, and apply quantization or distillation to the LLMs. The key indicator becomes "US$/action" or "US$/k-tokens" versus baseline, and it is desirable to maintain stable rates even with an increase in use.

***A second financial risk is tool sprawl** - each squad adopts a new SaaS or library every sprint. This lack of control increases indirect costs, hinders integration, and increases points of failure.*

The mitigation is an approved service catalog, with a "paved road" of official components, and TCO analysis before the "buy." Success is measured by reducing critical tools to a target number and by the coverage of the "paved road" in more than 90% of use cases.

Coupling and Evolution of the Data Layer

***Hard schema coupling occurs** when pipelines depend on specific columns and queries use SELECT *. Any change in the upstream can paralyze entire flows.*

Prevention requires versioned data contracts, a schema registry, and a linter that guarantees backward compatibility. Failures due to schema breakage should be below one incident per month.

***Another risk is LLM or API vendor lock-in:** if the provider increases the price or suffers latency degradation, you are left without an alternative.*

Use abstractions like LangChain or LlamaIndex, keep a fine-tuned open-source model of 7 B–13 B as a fallback, and enable multi-LLM routing (Azure, OpenAI, Anthropic). The goal is that at least 30% of requests pass through different providers, ensuring bargaining power.

Security and Privacy

he leak of PII via prompts is an immediate risk. *When personal data appears in the traces or logs, there is a LGPD/GDPR violation.*

Mitigate with a prompt firewall that detects PII by regex and real-time masking; also adopt row/column level security in streaming and vaults for secrets. The objective is to have zero incidents of leakage.

In parallel, prompt injection or jailbreak *allows users to induce the agent to execute unauthorized actions—for example, sending money or accessing confidential data.*

Implement policy guardrails, a whitelist of tools, and intention verification before any execution. Monitor the percentage of blocked requests and, even more importantly, ensure zero known false negatives.

Resilience and Availability

Event loss—stream gaps—arises when offsets become inconsistent or the lag exceeds the SLO. *This leads to incorrect calculations and subsequent decisions.*

The recommended practice is exactly-once configuration (Kafka + idempotence), "dead letters" queues for reprocessing, and automatic monitors that detect lag. A p95 lag below five seconds and gaps recovered in less than one minute are reasonable parameters.

Unpredictable latency *also harms agents that need to respond in real time.*

If the p99 exceeds the SLO at peak times, adopt hot caches (Redis or Pinot), circuit-breakers with graceful degradation, and local or edge inference for RAG. Aim to keep p99 latency under 500 ms.

Regulatory and Ethics

Unexplainable decisions put compliance at risk. *The absence of a complete trace of the flow prompt → reason → action makes audits unfeasible.*

Use immutable storage (WORM or ledger), record the RAG chain, and apply periodic auditing. Aim for 99% of actions with a complete trail.

Finally, algorithmic bias can generate statistically significant disparities between groups. *It is detected when fairness metrics deviate or complaints from affected users arise.*

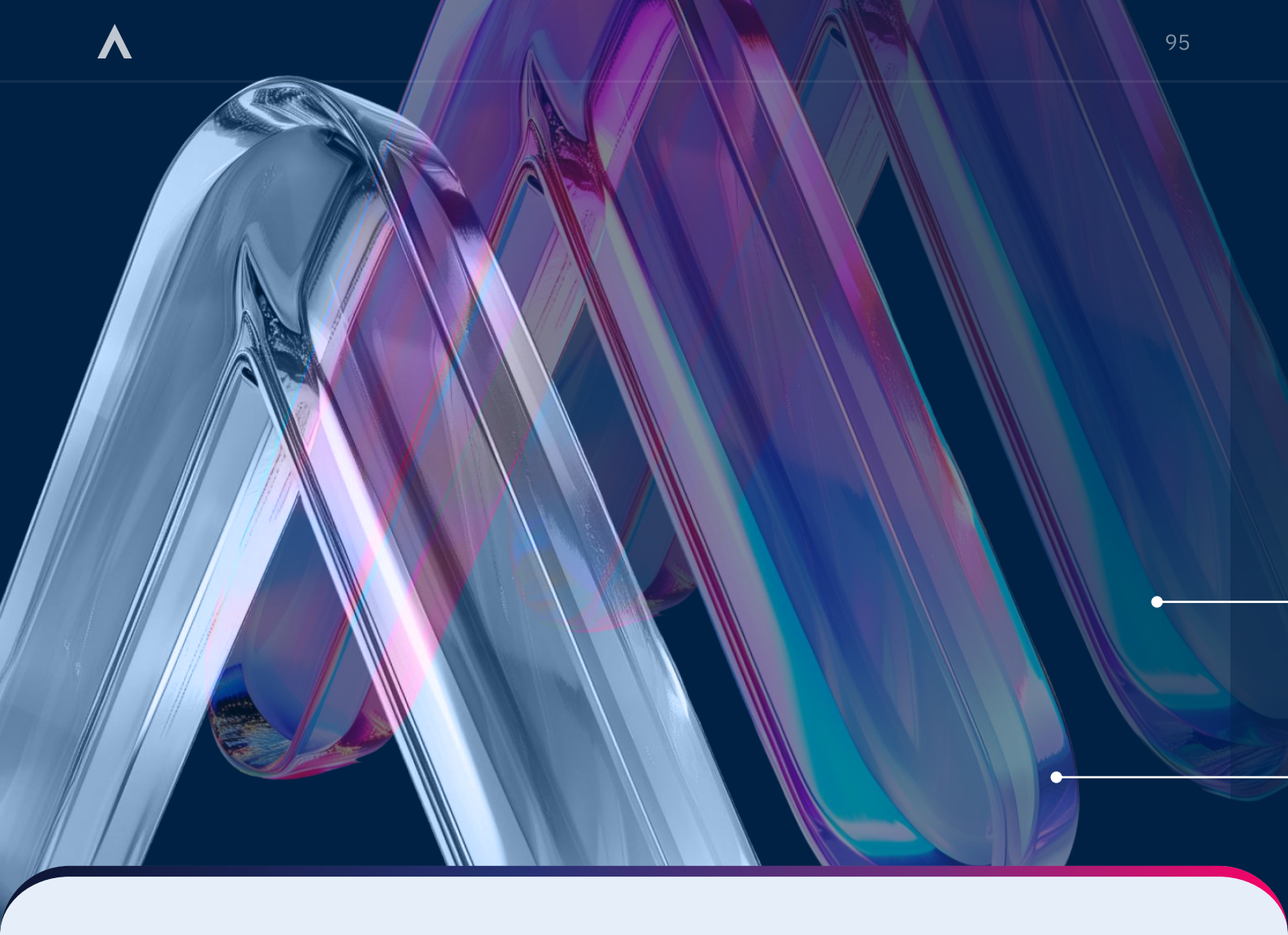
Mitigate with quarterly equity assessments, counterfactual tests in pre-production, and dataset balancing monitoring. The ideal is to keep disparities below the limit defined by governance (for example, a difference of less than "x %" with a p-value < 0.05).

Operationalizing Risk Management

*Integrate all these risks into a **living backlog radar**, with responsible parties, probability, impact, and review date.*

Describe controls as code (YAML, Terraform), associate GenAI SLOs—response quality, cost per token, freshness, and latency—with error budgets, and perform chaos engineering exercises to validate contingency plans (for example, LLM failure, stream delay, cost spike, or malicious injection). A multidisciplinary committee (data, security, legal, UX) should review logs, audits, and new use cases every quarter, deciding "go/no-go" for production.

When treated proactively, these risks cease to be obstacles and become governance levers: they ensure costs are under control, maintain regulatory compliance, and—above all—sustain the trust of users and stakeholders in the new generation of agentic platforms.



CHAPTER 8

Artefact’s Approach

Why Artefact?

> **Artefact is the ideal partner for companies that seek to transform their businesses through data and end-to-end artificial intelligence. With a specialized consulting model, we are a global company that connects business and technology, delivering innovative and personalized solutions that accelerate the adoption of data and AI.**

Our commitment is to unlock the value of data, democratizing its access and using it strategically to positively impact both people and organizations.

1 COMPLETE AND MULTIDISCIPLINARY EXPERTISE IN DATA AND AI

Artefact offers a complete and agnostic approach to technology. We operate in all stages of the data and AI journey, from strategy formulation to the execution and continuous management of solutions.

Our team is composed of more than 1,500 multidisciplinary specialists, including data analysts, data scientists, data engineers, and specialized consultants by sector. Each of these professionals brings a deep knowledge of data and AI, ensuring solutions that meet customer needs effectively and scalably.

Additionally, we are always at the forefront of technical evolutions, legal regulations, and the impact of data on business. **We are a data and AI pure player**, focused exclusively on delivering value through data-based solutions, ensuring that each project is aligned with the most advanced market trends and requirements.

2 PERSONALIZED AND SCALABLE SOLUTIONS FOR EACH NEED

We understand that each company has unique challenges and needs when it comes to data.

Therefore, Artefact offers personalized and scalable solutions that go beyond the simple implementation of technology. Our work model is completely customer-centric, and we ensure that each part of the implemented technology is adjusted to your specificities and is in your own infrastructure, without leaving your systems. This ensures total security and control of the data.

OUR RANGE OF SERVICES INCLUDES:

- **Data Strategy and Governance:** Development of data strategies aligned with business objectives and implementation of good governance practices.
- **Data and AI Factory:** Construction of data platforms and implementation of AI solutions that drive commercial operations, marketing, and innovation.
- **System Integrations:** We facilitate the integration of systems and tools, ensuring that data flows efficiently between different platforms.
- **Predictive Analysis and GenAI:** We use AI/ML and GenAI to optimize decision-making, from demand forecasting to increasing personalization and product innovation.

Additionally, our experience extends to sectors such as **Data-Driven Sales and Marketing**, with a focus on the digital transformation of commercial practices, using data for more assertive decisions and personalized strategies.

3 SUSTAINABLE TRANSFORMATION WITH KNOWLEDGE TRANSFER AND OWN INFRASTRUCTURE

One of Artefact's great differentiators is our philosophy of knowledge transfer and IP.

We work collaboratively with our clients' internal teams, creating hybrid teams that ensure the ownership of the implemented solutions and technologies. With this, we ensure that companies not only implement digital transformation but also become capable of managing and innovating autonomously in the future.

We are committed to creating a data culture in organizations, through specific adoption programs such as **data schools, bootcamps, and hackathons**, developing skills in internal teams to maintain continuous innovation.

Artefact is more than a consultancy; we are a strategic partner in the development of solutions that go beyond technology.

We offer specialized consulting by sector, personalized software engineering, and practical experience, making us the ideal choice for companies that seek to make the most of the value of data and AI.

What makes us unique:

- ✓ **Technological Agnosticism:** We implement solutions in your infrastructure, without leaving your systems.
- ✓ **Multidisciplinary Expertise:** A team of more than 1,500 specialists combining innovation, data science, engineering, and consulting.
- ✓ **Responsible AI Certification:** Pioneers in the Label IA certification, ensuring ethical and responsible practices in the use of AI.
- ✓ **Complete Transformation:** From strategy to execution, with a focus on sustainability and training of internal teams.

Artefact is your partner to not only transform your data into valuable assets but also to ensure that this transformation is lasting and sustainable, connecting your business to the next generation of data and AI technology.

Businesses we've already impacted with GenAI and Agentic AI

DANONE

Demand and Supply Forecasting

With Artefact's forecasting solution, we increased accuracy and reduced the time required by internal teams. In just three months, we achieved significant financial gains.

HEINEKEN

Stockout Forecasting

With Artefact's stockout forecasting, we mitigated risks, increased efficiency, and safeguarded revenue across the entire supply chain — all in just five days.

Some of our clients

HEINEKEN, L'ORÉAL, DANONE ONE PLANET. ONE HEALTH, orange, Carrefour, LVMH, FORTENOVA GROUP, Unilever, MICHELIN, SAMSUNG, BAYER, Johnson & Johnson, Opella., sanofi, COPA energia, suzano, ArcelorMittal, Santander, BNP PARIBAS, SOCIETE GENERALE, adyen, LCL Your Data Center, ANBIMA, J.P.Morgan, AXA

ARTEFACT
AI IS ABOUT PEOPLE

WE OFFER **END-TO-END** DATA & AI SERVICES

Strategy & Transformation

- Data & AI Strategy
- Data Maturity Assessment
- Hackathons
- GenAI Academy
- Data & AI organization
- Corporate Training
- Data & AI Days
- Artefact AI Summits

AI Acceleration

- AI & Gen AI Factory
- Data & AI for Operations
- AI for Customer Care
- Data & AI for Private Equity

Data Foundations & BI

- Data Governance & Management
- Data – New BI – Self Business Intelligence
- Data for Sustainability

IT & Data Plataforms

- Data-Centric IT
- Cloud Services
- Tech-Agnostic Solutions
- Smarter Decision-Making
- Optimized IT Operations

Marketing Data & Digital

- Consumer Data Environment
- Measurements (MROI) & Insights
- Data-driven Sales
- Data Valorization & Category Management
- Marketing Analytics
- GMP Certified Reseller

Expertise by Industry:

FMCG | RETAIL & E-COMMERCE | LUXURY & COSMETICS | HEALTHCARE | BANKING & INSURANCE | TELECOMMUNICATIONS | SPORTS & ENTERTAINMENT | TRAVEL & TOURISM | PUBLIC SECTOR & GOVERNMENT | REAL ESTATE | MANUFACTURING & UTILITIES

CHAPTER 9

Conclusion

Data Platforms: O Pilar da Inovação e Crescimento Sustentável

Throughout this e-book, we explored the evolution of data architecture and the impact of data platforms on the digital transformation of companies. We observed how the transition from the traditional data model to modern platforms is redefining the way organizations manage and leverage their data for strategic decisions. The challenges faced by companies, from the complexity of legacy systems to the integration of new technologies, are evident, but they are also surmountable with the right approach.

> **The future of data management has already begun to take shape, and companies that adopt modern data platforms, such as Data Lakehouses, Data Mesh, and advanced AI/ML solutions, are not only overcoming these challenges but also creating new opportunities for innovation and sustainable growth.**
>
> These solutions not only allow companies to centralize, integrate, and transform their data efficiently but also enable personalization at scale, bringing tangible results such as greater operational efficiency, more precise forecasts, and faster decisions.

data platforms will be better positioned to make more informed decisions, reduce costs, and generate a positive impact on relationships with customers, partners, and stakeholders. Data architecture is, without a doubt, a central pillar for the businesses of the future—where artificial intelligence and data go hand in hand to accelerate digital transformation.

The adoption of these platforms requires a clear vision, strategic planning, and rigorous execution.

But, as we saw in innovative sectors, the benefits of adopting the correct data architecture are evident: efficiency, continuous innovation, and a competitive differentiator that transforms data into one of an organization's greatest assets.

If there is one certainty in this transformation scenario, it is that companies that start to implement and scale their data platforms now will be at the forefront of the digital revolution, shaping the future of their industry and becoming leaders in an increasingly data-driven market.

The path to success is clear, and the key is to transform the way companies treat and use their data to create a real and lasting impact.

Links & References

Research & Consulting

Atlan: *"Gartner Data Governance Maturity Model: What It Is, How It Works."* Acesso em: 11 de julho de 2025. Disponível em: https://atlan.com/know/gartner/data-governance-maturity-model/

Technical Articles & Blogs

Control Global: *"The power of generative AI and advanced analytics."* Acesso em: 17 de julho de 2025.
Disponível em: https://www.controlglobal.com/manage/data-analytics/article/33015567/the-power-of-generative-ai-and-advanced-analytics 2

Research & Consulting

Databricks: *"Databricks Unified Data Analytics Platform Can Generate 417 Percent Return on Investment, According to TEI Study."* Acesso em: 11 de julho de 2025.
Disponível em: https://www.databricks.com/company/newsroom/press-releases/databricks-unified-data-analytics-platform-can-generate-417-percent-return-on-investment-according-to-tei-study 3

Technical Articles & Blogs

Dataguard: *"What is Zero Trust Architecture, and is it a must in your security?"* Acesso em: 17 de julho de 2025.
Disponível em: https://www.dataguard.com/blog/what-is-zero-trust-architecture-is-it-a-must-in-security/ 4

Technical Articles & Blogs

FME: *"AWS Cost Management in Data Lake Projects."* Acesso em: 11 de julho de 2025.
Disponível em: https://en.fme.de/blog/aws-cost-management-in-data-lake-projects/ 5

Article

IBM: *"What is advanced analytics?"* Acesso em: 11 de julho de 2025.
Disponível em: https://www.ibm.com/think/topics/advanced-analytics 6

Article

IBM: *"Why is data architecture important?"* Acesso em: 10 de julho de 2025.
Disponível em: https://www.ibm.com/think/topics/data-architecture 7

Research & Consulting

McKinsey & Company: *"Reducing data costs without jeopardizing growth."* Acesso em: 11 de julho de 2025.
Disponível em: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/reducing-data-costs-without-jeopardizing-growth 8

Academic Studies & Publications

ScienceDirect: *"Data governance: A conceptual framework, structured review, and research agenda."* Acesso em: 15 de julho de 2025. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0268401219300787 9

Technical Articles & Blogs

The Seattle Data Guy: *"Data Warehouses vs. Operational Data Stores vs. Data Lakes: How to Store Your Data for Analytics."* Acesso em: 11 de julho de 2025. Disponível em: https://www.theseattledataguy.com/data-warehouses-vs-operational-data-stores-vs-data-lakes-how-to-store-your-data-for-analytics/#page-content 10

Academic Studies & Publications

Tandfonline: "*The contribution of GenAI to business analytics.*" Acesso em: 17 de julho de 2025. Disponível em: https://www.tandfonline.com/doi/full/10.1080/2573234X.2024.2435835#d1e40311

Technical Articles & Blogs

Thoughtspot: *"What is advanced analytics? Definition, benefits, and use cases."* Acesso em: 11 de julho de 2025. Disponível em: https://www.thoughtspot.com/data-trends/analytics/advanced-analytics 12

Glossary

Data Lake

A repository for storing large volumes of raw data, including unstructured formats (such as videos, audios, and texts). It offers flexibility but requires good governance not to become a "data swamp."

Data Lakehouse

A hybrid architecture that combines the flexibility of the Data Lake for raw data with the structure and performance of the Data Warehouse for quick analyses.

Data Mart

A subset of a Data Warehouse, focused on a specific department or business area. It allows for faster and more targeted analyses, without the complexity of the central system.

Data Mesh

A decentralized approach where different business areas manage their own data as "products." It promotes more autonomy and agility, eliminating bottlenecks from centralized teams.

Data Vault

A data architecture that prioritizes traceability and governance. Its structure allows data to be stored in a flexible and auditable way, making it ideal for regulated sectors.

Data Warehouse (DW)

A centralized system for storing structured data from different sources. Ideal for historical reports and analyses, it serves as the company's "single source of truth."

GenAI (Generative Artificial Intelligence)

A branch of AI focused on creating new content (text, images, code) based on existing data.

LLM (Large Language Model)

Language models, such as GPT, that are the basis of GenAI. They understand and generate text in natural language.

ODS (Operational Data Store)

A repository that integrates real-time operational data from different systems. It is used to support quick day-to-day decisions.

RAG (Retrieval-Augmented Generation):

A technique that uses LLMs to generate responses, but based on a previous search in reliable company data. This makes the responses more precise and avoids the so-called "hallucinations."