

# Data Observability: The Key to Trustworthy Data Pipelines and Analytics

In today's data-driven world, organizations are accumulating vast amounts of data at an unprecedented pace. This data serves as the lifeblood of decision-making processes, fueling analytics, machine learning and business intelligence initiatives. However, the value of data is directly proportional to its availability and quality. Pipeline complexity, often resulting from intricate data silos, can hinder data accessibility and make it challenging to maintain high-quality data. Additionally, stringent data privacy regulations necessitate careful handling of data, raising the stakes for ensuring quality while safeguarding personal information. Successful data initiatives also depend on collaboration, as complex data ecosystems require multifaceted efforts to maintain data availability and quality.

Data availability is essential for all data-driven businesses. It is the foundation upon which modern business operations are built. When data is unavailable, its value is lost. Critical business operations can be disrupted, leading to missed opportunities, workflow disruptions and financial loss. Data unavailability can manifest in various ways, such as system downtime, data silos and technical glitches. Even a moment of data inaccessibility can have far-reaching consequences in the fast-paced world of business. It is therefore imperative to safeguard it. This requires robust infrastructure, data management practices and a commitment to proactive monitoring and swift issue resolution.

Data quality is equally vital. It is an essential cornerstone for both analytics and business processes. In analytics, the integrity of insights and decisions heavily relies on the accuracy and reliability of data. High-quality data ensures that patterns, trends and correlations derived through analysis are trustworthy, enabling sound decision-making and predictions. In the context of business processes, data quality is vital for operational efficiency and informed decision-making. Reliable data supports day-to-day operations, from inventory management to customer interactions, enhancing process efficiency and reducing errors.

## Key Benefits

- Speed up time to resolution of data quality issues
- Enhance the bottom line from optimized pipelines and resource utilization
- Comply with regulations using a detailed view of data lineage
- Improve predictions and business outcomes with trusted, timely data

This is where a data observability solution comes into play, supported by robust data quality measures. Companies must have strong capabilities in data observability and discoverability to make data accessible across the organization (McKinsey, "Bringing Data Insights to Cloud,") 2023).<sup>1</sup>

## Stay Ahead of Data Issues with Intelligent Data Observability

Data observability empowers organizations to maintain data integrity, security and compliance while maximizing the value of their data assets. This capability helps to track data consumption, adhere to policies and regulations, ensure data health and quality and optimize data pipelines. Data observability can be viewed through three distinct lenses: business, data and pipelines. Each of these lenses offers a unique perspective on the monitoring and enhancement of data-related aspects:

- **Business:** Concentrates on how data is consumed, protected and compliant with policies and regulations
- **Data:** Addresses the health of data, identifying the impact of any issue and taking prompt preventative or remedial action
- **Pipelines:** Focuses on optimizing the availability, performance and capacity of data pipelines

A comprehensive data observability solution should encompass these elements to ensure data integrity, security, compliance and operational excellence.

## Introducing Cloud Data Management

The **Informatica Intelligent Data Management Cloud™** (IDMC) offers a comprehensive set of services to help establish proactive data observability, emphasizing anticipation over reaction to potential data issues. The platform provides a unified solution encompassing data profiling, **data quality, data lineage, data governance, data sharing** and pipeline optimization. This common platform facilitates a flexible yet unified approach to data observability with tailored insights for data engineers, IT operations, chief data and analytics officers and data analysts.

The **IDMC Cloud Data Quality** service leverages **AI-powered CLAIRE®** to identify issues and anomalies in the data, and checks whether data meets predefined quality standards on an ongoing basis. Once issues and anomalies are identified key stakeholders can be alerted. These alerts serve as early warnings, prompting timely remediation actions. The insights derived from continuous monitoring and remediation efforts contribute to enhancing data systems, improving data quality and making the organization more resilient against data-related challenges.

<sup>1</sup> <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/bringing-data-platforms-to-cloud>

The **IDMC Data Governance** and **Catalog** service provides the overarching framework for how an organization manages its data. This service encompasses policies, procedures and practices to ensure that data is collected, stored, processed and used effectively and responsibly. The data lineage capability within this IDMC service provides the visual representation of data's journey from source to destination. Data lineage shows how data is acquired, processed and transformed as it moves through an organization's systems. This visualization aids organizations in understanding the potential impact of data issues on downstream applications and processes.

The **IDMC Cloud Data Marketplace** service offers a centralized, self-service data exchange for publishing curated data products. It enables data consumers to access, consume and collaborate on the available datasets. Integrated Cloud Data Quality provides observability across data assets contributed by distributed data domain owners. This helps ensure that data products meet centralized data quality standards and are available to data consumers with authorized access.

The **IDMC Operational Insights** service provides a holistic approach to managing data pipeline resources and performance. It encompasses cost optimization, resource allocation, bottleneck identification, streamlined workflows, performance enhancements and data lifecycle management. This can ensure an organization's data pipelines are cost-effective, highly efficient and well-aligned with their business goals.

Informatica's comprehensive approach to data observability encompasses a range of key features and capabilities, which include:

- Providing visibility into data health to quickly resolve issues
- Observing data flows for critical jobs
- Optimizing availability, performance and capacity
- Improving data agility and compliance
- Tracking and optimizing resource consumption with FinOps

## Key Capabilities

### Providing Visibility into Data Health to Quickly Resolve Issues

Staying ahead of data quality issues and anomalies is an essential capability that enables organizations to swiftly identify and address potential data challenges before they escalate. This proactive approach not only ensures that data remains accurate, consistent and reliable but also minimizes disruptions to business operations. By continuously monitoring data streams and processes, organizations can spot anomalies, outliers or deviations from established business rules, allowing for immediate alerts and intervention. This approach streamlines the remediation process, reducing downtime and preventing the recurrence of similar issues in the future. Detecting data quality issues beforehand empowers organizations to maintain the integrity of their data, make informed decisions and bolster operational efficiency which contributes to the overall success of data-driven initiatives.

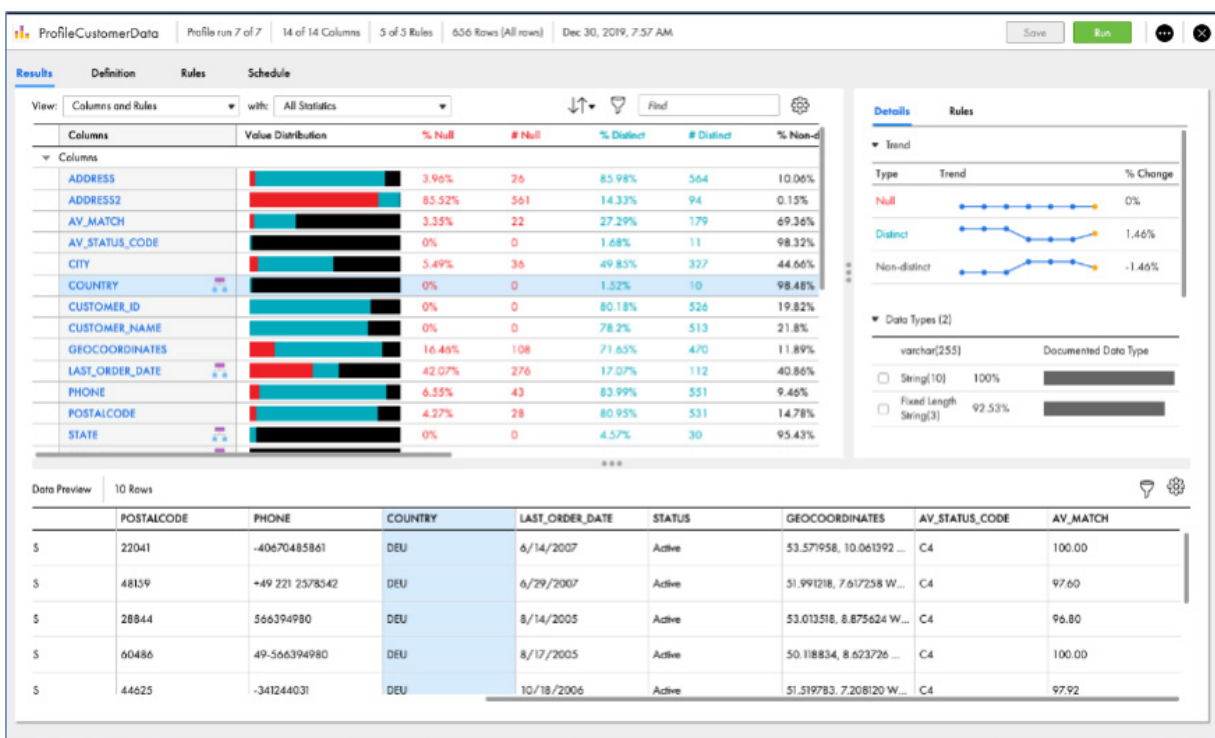


Fig 1. Data Quality insights: Ensure that data is complete, accurate and consistent and pinpoint the source of issues and the potential impact.

## Observing Data Flow and Automated Resolution for Critical Jobs

Continuous insights for rapid issue detection and automated resolution are at the core of efficient data management. With real-time monitoring and analysis, organizations can identify data issues as they occur, offering timely intervention to prevent disruptions. The ability to gather continuous insights from data streams enables quick detection of anomalies, inconsistencies and deviations. This ensures that data quality remains high. Furthermore, with automated resolution mechanisms in place, organizations can streamline the process of fixing data problems, reducing manual intervention and saving valuable time and resources. This approach not only enhances data reliability but also paves the way for seamless, uninterrupted business processes, facilitating agile decision-making and data-driven excellence.

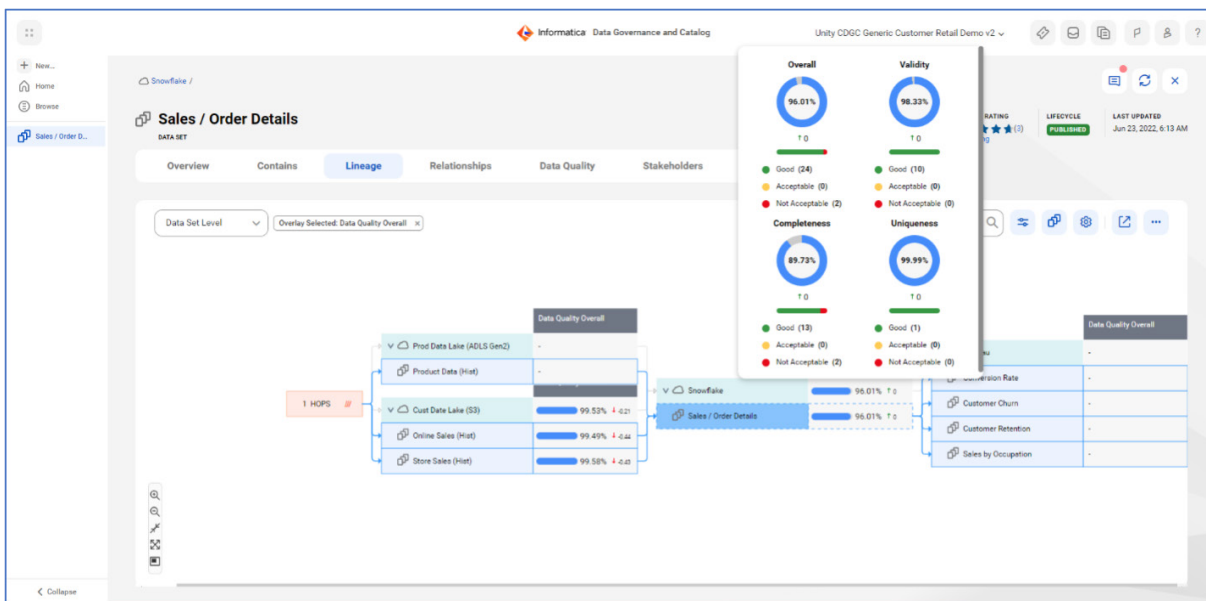


Fig 2. Data Lineage and Quality: Gain additional understanding of your data by displaying graphic overlays, including data quality scores, and view relevant information alongside data lineage.

## Optimizing Availability, Performance and Capacity

The optimization of data availability, performance and capacity is a crucial objective for organizations seeking to maximize the value of their data resources. This process involves fine-tuning data systems to ensure that data is consistently accessible, responsive and scalable. By achieving optimal data availability, organizations minimize the risk of downtime and maintain uninterrupted access to critical information.



Performance optimization guarantees that data processes and analytics run efficiently, enabling faster insights and more responsive applications. Capacity optimization ensures that data systems can scale with growing data volumes, supporting the ever-expanding requirements of modern businesses. This comprehensive approach not only enhances the robustness of data infrastructure but also empowers organizations to harness the full potential of their data for strategic decision-making and operational excellence.

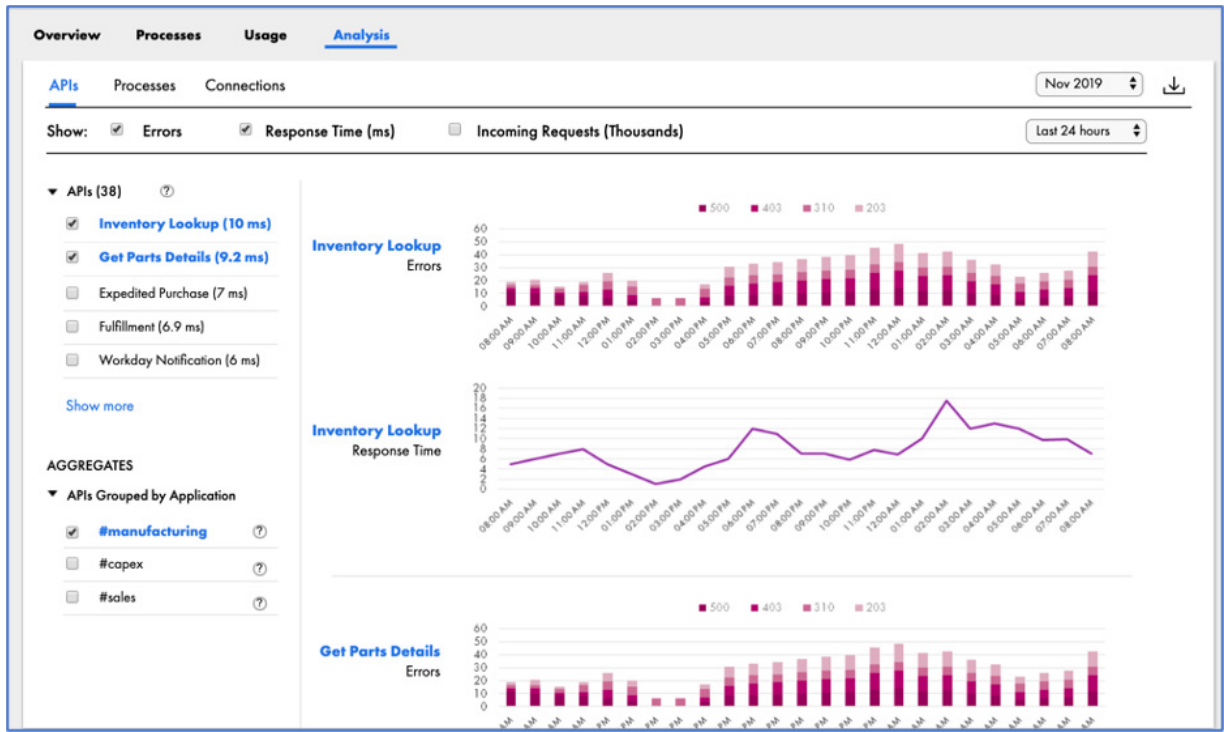


Fig 3. Operational Insights: Autonomously detect and rectify issues to minimize downtime and ensure continuous data flow.

### Improving Data Agility and Compliance

Gaining a deeper understanding of how data is being accessed and used is an essential aspect of effective data management. It involves tracking and analyzing data usage patterns, which can unveil valuable insights about how data assets are leveraged within an organization. By delving into this information, organizations can identify which data is most critical, who the primary users are, and how data contributes to various business processes.

This understanding allows for more informed decisions regarding data storage, access permissions and data lifecycle management. Moreover, it can aid in optimizing data resources, ensuring that data is available to those who need it most, while also safeguarding sensitive information. In essence, gaining deeper insights into data usage is a cornerstone of data governance, enhancing data-driven decision-making and aligning data management strategies with the specific needs and objectives of the organization.

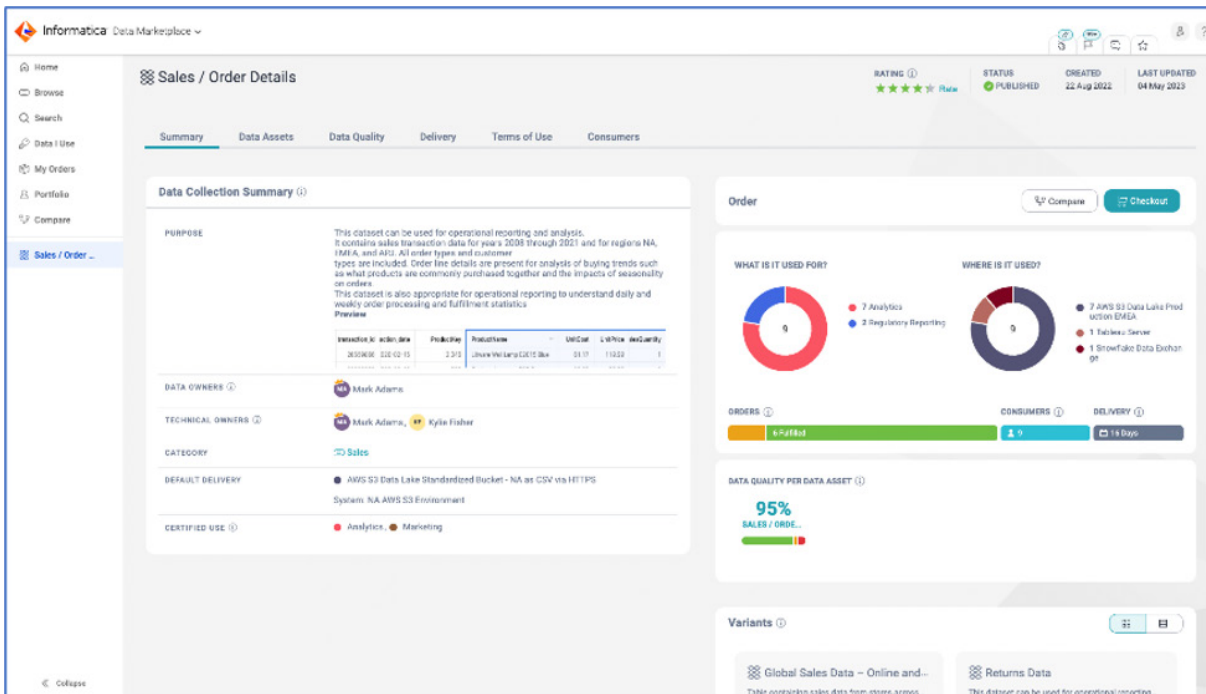


Fig 4. Data Marketplace: Identify who owns the data, who changes the data, who queries the data, and how often the user uses the data.

## Tracking and Optimizing Resource Consumption with FinOps

Implementing FinOps (Financial Operations) in data pipelines harmonizes financial and operational aspects of data management, enhancing efficiency and effectiveness. This involves meticulous cost monitoring, resource optimization and performance improvements in data pipelines. Implementing FinOps principles helps organizations allocate and track data-related costs, promoting judicious resource usage for cost-effective operations. By fine-tuning data pipelines, identifying bottlenecks, optimizing resource allocation and streamlining workflows, organizations can ensure peak efficiency. This aids in delivering high-quality data for informed decision-making and driving operational excellence.

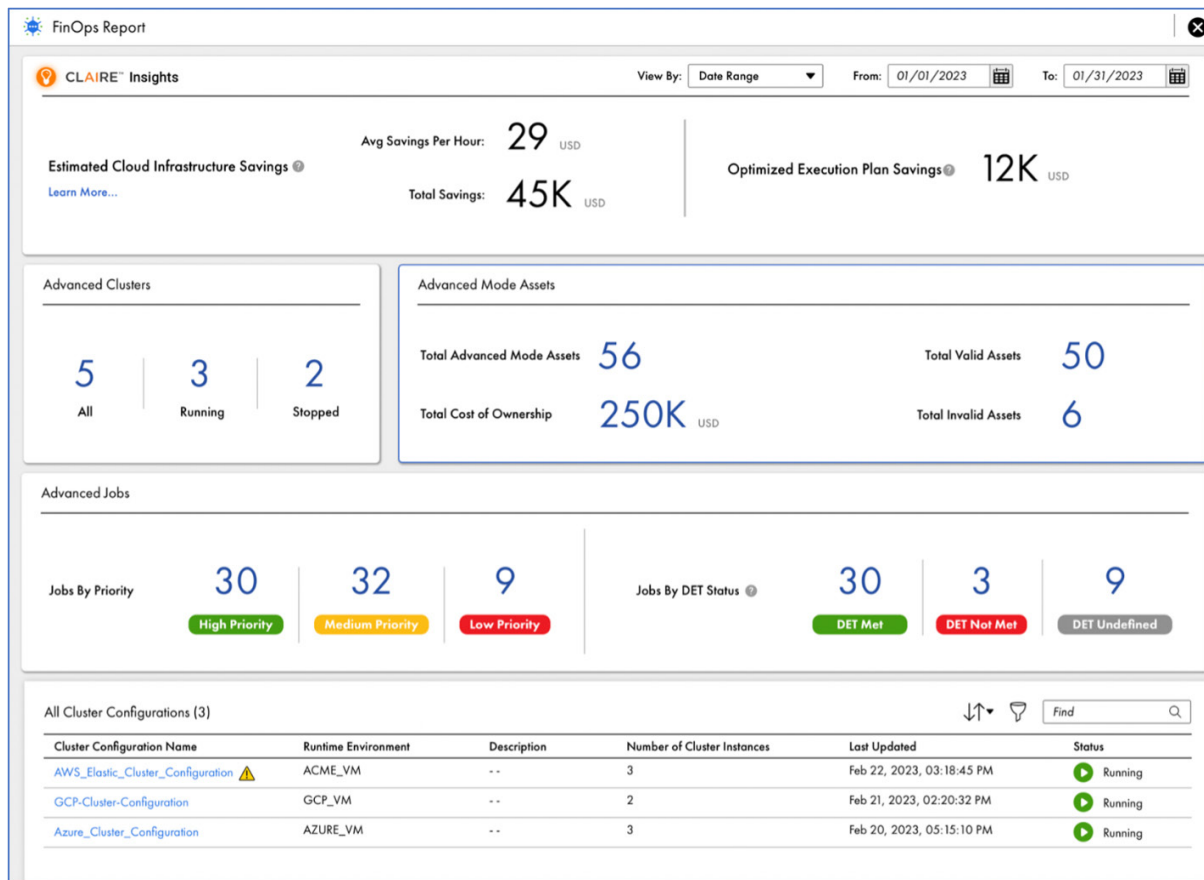


Fig 3. Operational Insights: Autonomously detect and rectify issues to minimize downtime and ensure continuous data flow.

## Key Benefits

### Speed Up Time to Resolution of Data Quality Issues

The ability to quickly identify and understand data issues is essential for data observability, as it contributes to data reliability, operational efficiency and data-driven decision-making. Resolving data challenges proactively allows organizations to maintain the integrity of their data systems, ensuring that data remains a valuable and trustworthy asset for their data-driven initiatives.



## Enhance the Bottom Line from Optimized Pipelines and Resource Utilization

Data observability helps organizations improve their bottom line by providing visibility into data systems, optimizing data pipeline and improving resource utilization. This can lead to identifying and eliminating inefficiencies, reducing costs and enhancing operational efficiency, which all contribute to improved financial outcomes and a competitive edge.

## Comply with Regulations Using a Detailed View of Data Lineage

Delivering a detailed view of data lineage within the context of data observability ensures the proper usage of data but also supports compliance with data privacy and regulatory requirements. Moreover, it contributes to data governance impact analysis and overall data reliability.

## Improve Predictions and Business Outcomes with Trusted, Timely Data

By ensuring that data is delivered promptly and maintained at a high level of quality and trustworthiness, organizations can make faster, more accurate decisions, respond to changing circumstances and ultimately achieve better business results.

### Where data & AI come to



#### Worldwide Headquarters

2100 Seaport Blvd., Redwood City, CA 94063, USA Phone: 650.385.5000, Toll-free in the US: 1.800.653.3871

Informatica (NYSE: INFA) brings data and AI to life by empowering businesses to realize the transformative power of their most critical assets. When properly unlocked, data becomes a living and trusted resource that is democratized across your organization, turning chaos into clarity. Through the Informatica Intelligent Data Management Cloud™, companies are breathing life into their data to drive bigger ideas, create improved processes, and reduce costs. Powered by CLAIRÉ®, our AI engine, it's the only cloud dedicated to managing data of any type, pattern, complexity, or workload across any location — all on a single platform.

IN17-4695-0124

© Copyright Informatica LLC 2024. Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and other countries. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners. The information in this documentation is subject to change without notice and provided "AS IS" without warranty of any kind, express or implied.

Data observability is crucial for modern data management and benefits data engineers by helping them identify and fix pipeline bottlenecks, reducing downtime and improving data flow. This helps to enable many other data users to benefit throughout the organization.

Data scientists can gain confidence in their analyses, ensuring accurate data use. Business analysts are able to make informed decisions based on reliable data, improving organizational efficiency. Chief data officers (CDOs) benefit from greater compliance, leveraging data observability to ensure data use aligns with regulations and privacy requirements, reducing the risk of non-compliance and data breaches.

## Next Steps

Learn more about how Informatica's **intelligent data observability tools** intelligent data observability tools can help you connect data consumers with timely, trusted data.