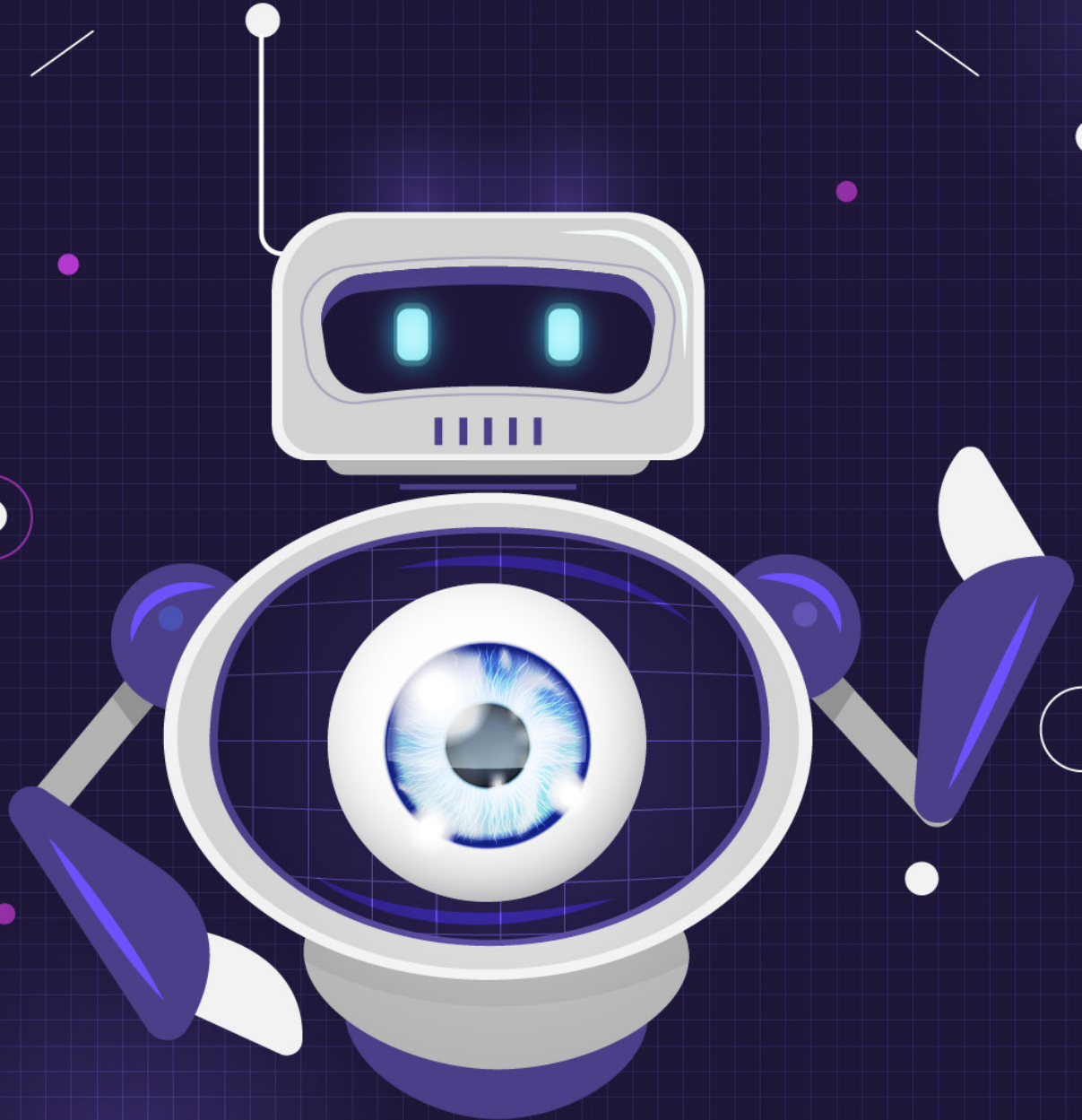


DATA OBSERVABILITY



An Enterprise Data Management Imperative Gaining
Attention in Boardrooms

WWW.COURSE5I.COM

Authored By

Devendra Desai

Principal Solution Consultant

Shreyas Sundaresh

Client Success and Growth

Why Read this Paper:

With fast-expanding use cases of enterprise data to understand a business and its customers and competitors in real-time, organizations are investing significantly in setting up their data strategy.

Yet, system breakdowns and ETL disruptions due to unavailable, inaccessible, or inaccurate data and erroneous metrics are common. Further, the lack of real-time visibility on system metrics and data outputs prevents organizations from truly leveraging their underlying data to improve the functionality of the data. System lags and disruptions continue to impact decision-making, business performance, customer service levels, and the reputation of an organization.

This is where Data Observability comes in.

Data Observability **enables real-time monitoring and understanding of a data system's behavior and performance**, looking at aggregated metrics and insights about the data system and its processes to identify issues, patterns, trends, and causes to eliminate or quickly resolve malfunction. Importantly, Data Observability provides visibility into the **historical context of data**, which includes the origination of data from the source system, data flow within the ecosystem of downstream systems and applications, and the processing and transformation (ETL) that the data is subject to. This visibility becomes critical for quality assurance, governance, compliance, root cause, and impact analysis.

Who is it for?

- CXO, Business, and Product teams
- Data, Analytics, MLOps, and Cloud Operations teams
- Managed Services stakeholders:

Executive Summary:

This whitepaper introduces Observability as a critical component of overall Enterprise Data Management. It explores the current challenges and vulnerabilities of the broadly adopted practices within Enterprise Data Management, and how Data Observability comes to the rescue. The paper briefly touches on how Course5 Intelligence creates business impact in organizations through Data Observability, and highlights some of the challenges and best practices to be followed for successful implementation and adoption of Data Observability.

This paper explores the making of Observability a reality for the data and analytics ecosystem. Through this exploration, the paper aims to answer the following key questions:

How do organizations leverage Observability and "business domain" together to enable automated Root-cause analysis across multivariate business and operation signals?

What challenges and complexities arise when implementing Observability that predicts the failures at a transaction scale in the context of distributed systems, microservices architectures, and emerging technologies?

What foundational components constitute Observability for implementing much more rapidly than ever before?



What is the Observability Scope that includes emerging trends and advanced approaches like "AI/ML" that enable Intelligent alerting, anomaly detection, etc., for shaping the trajectory of this field?

What are the framework and best practices for successfully implementing and adopting data Observability that enables self healing capability?

**Data Observability:
A Key Tenet to Enterprise Data Management**

What is Observability?

Data Observability, which is an organization's ability to understand the health and quality of data within its ecosystem, has gone from being a blog post to a listing on the Gartner Hype Cycle for Emerging Technologies due to its focus on:

- Reducing the number of data incidents
- Detecting potential data incidents earlier
- Resolving data incidents faster
- Preventing data incidents

Observability represents the capacity to gain insights into a system's internal state and functionality through an analysis of its external outputs. This concept holds great significance in the business landscape of complex systems, where it plays a pivotal role in ensuring reliability, performance, and security:

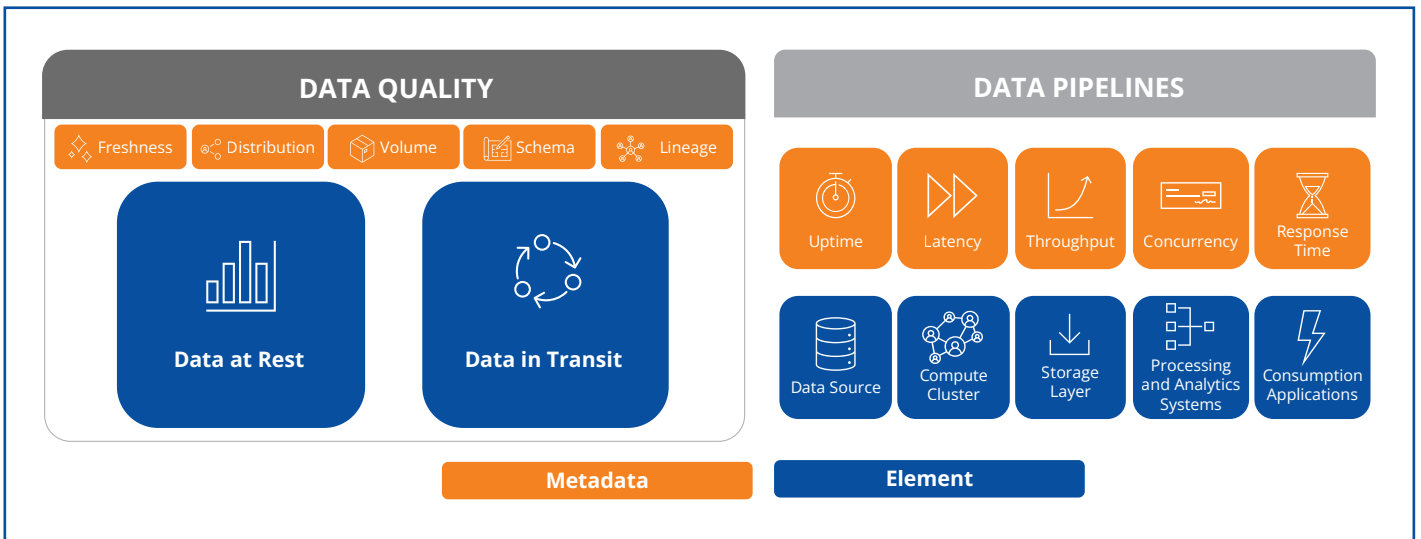
- **Reliability Enhancement:** Reliability Enhancement: Observability serves as a proactive tool for identifying and rectifying issues before they disrupt vital business processes. Monitoring a system's external behavior allows for the early detection of anomalies, thereby minimizing downtime and bolstering overall system reliability.
- **Performance Optimization:** Observability lends itself to the detection of bottlenecks within processes, enabling systematic improvements for optimized performance. This heightened visibility into system dynamics facilitates superior resource management, culminating in an enhanced user experience.
- **Security Fortification:** Real-time threat detection and response constitute a critical facet of observability. By scrutinizing external outputs, observability can swiftly identify and counteract security breaches, thwarting unauthorized access, data breaches, and other security threats.

In essence, observability is defined by its ability to answer questions regarding a system's state and behaviors. This is achieved through the systematic collection and analysis of various forms of system data, including logs, metrics, and traces.

In the context of data and analytics (D&A), Data Observability finds application across an expansive array of domains. It encompasses data quality assurance, data drift monitoring, data pipeline management, data orchestration, operational oversight, analytics, Machine Learning model deployment, system performance, and infrastructure management. Additionally, it encompasses AI-driven optimization and recommendations aimed at upholding data quality, reliability, performance, and security.

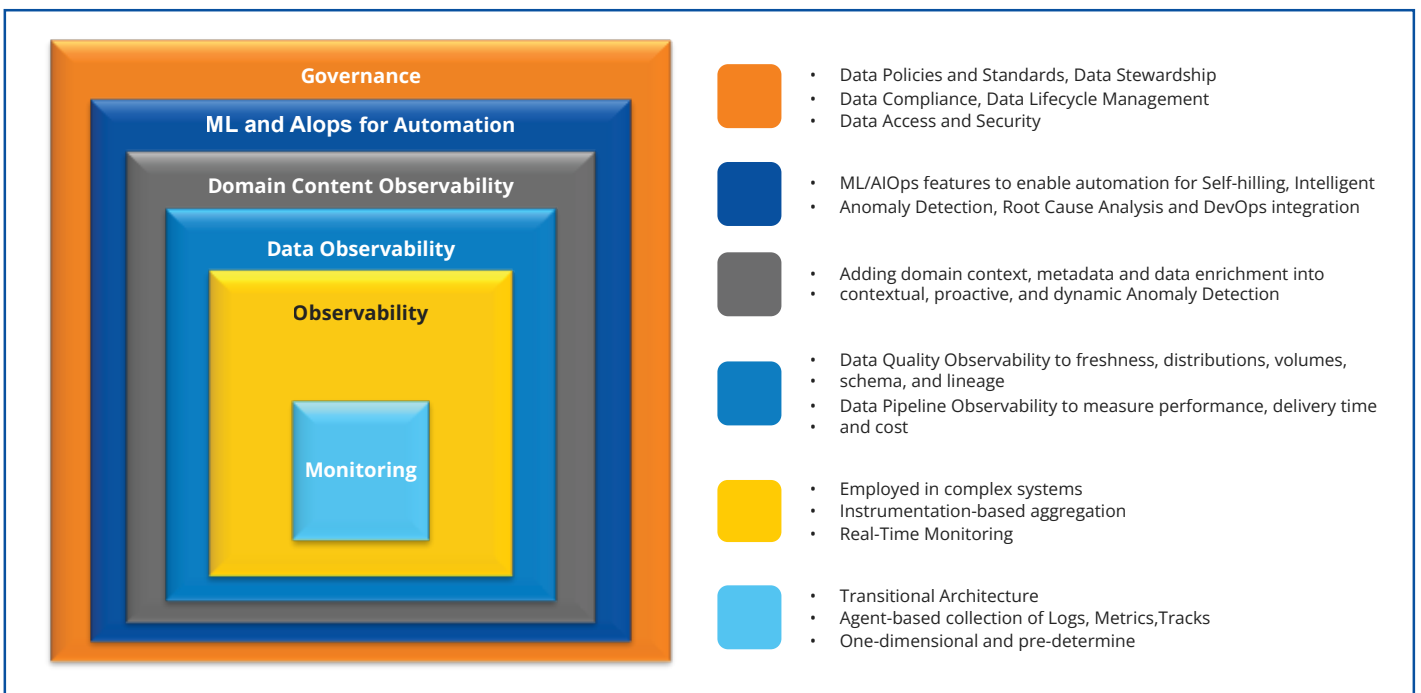
While Observability (Software) and Data Observability share a common foundational premise, they diverge in their areas of focus within a system and its data. While DevOps employs Observability in the software domain, data teams adopt a rigorous approach to data with Data Observability. The holistic approach to data platforms unfolds along two key dimensions:

- **Data Quality Observability:** This facet concentrates on evaluating the quality and timeliness of data. It involves the monitoring of data in transit and at rest to gauge attributes such as freshness, distribution, volume, schema adherence, and lineage.
- **Data Pipeline Observability:** The emphasis here is on scrutinizing the quality and performance of data pipelines, encompassing the underlying infrastructure that supports them. This dimension involves the observation of data sources, computational clusters, storage layers, processing and analytics systems, and consumption applications. Key metrics include uptime, latency, throughput, concurrency, delivery time, cost of ownership, and response time.



Is it different from Monitoring and Governance?

Observability and monitoring are distinct yet complementary concepts. Monitoring serves to identify issues when they occur, while observability delves into the reasons behind these occurrences. Monitoring is a subset of observability, playing a vital role in its framework.



Monitoring vs. Observability vs. Governance

Monitoring (APM): Monitoring primarily focuses on tracking specific predefined metrics and data points from a system or application to assess its health and performance. It operates reactively by triggering alerts or notifications when predefined thresholds are breached or specific events transpire. Monitoring excels at identifying known problems, resource utilization tracking, ensuring system availability, and adhering to service-level agreements (SLAs).

Data Observability: Data observability extends monitoring to encompass the data's health, quality, and performance as it traverses various systems and processes. Unlike monitoring, it relies less on predefined metrics, encompassing a wide range of data types such as logs, traces, and event streams for a holistic system view. Observability excels in troubleshooting complex issues, uncovering root causes, and providing insights during unforeseen scenarios, especially in dynamic environments like microservices and the cloud.

With the advent of cloud virtualization and increased traffic complexity, traditional monitoring systems struggle to correlate diverse data sources effectively. This leads to challenges in root cause analysis and issue resolution, causing frustration among support teams and end users. Observability, a novel paradigm for enhancing visibility in complex distributed systems, addresses these shortcomings.

On the other hand, Data observability and governance are related concepts within data management, but they address different aspects of data quality, monitoring, and control.

Data Governance: Data governance encompasses practices, processes, and policies that ensure data quality, availability, security, and compliance across an organization. It involves planning, execution, and control of data-related activities to manage data effectively, align it with business objectives, and ensure its trustworthiness.

While data observability focuses on monitoring data pipelines and anomaly detection, data governance has a broader scope, encompassing activities related to data quality, compliance, security, and overall management. Data observability can complement data governance by providing real-time insights into data health and performance, aiding organizations in maintaining high-quality, trustworthy data.

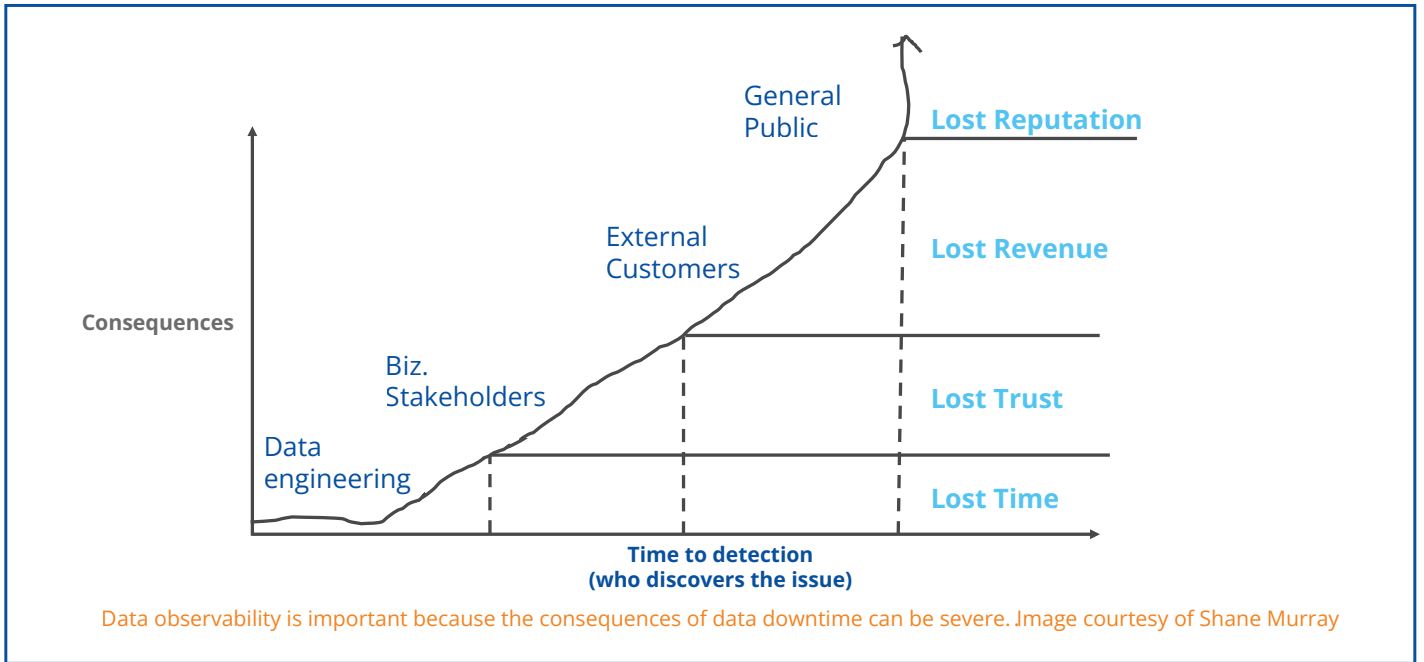
Let's explore how they are related yet distinct:

Key Aspects	Monitoring (APM)	Data Observability	Governance
Transitional Architecture	Y		
Agent-based Collection of Logs, Metrics and Tracks	Y		
One-dimensional and Pre-determine	Y		
Employed in Complex Systems with heterogeneous deployment architecture and distributed systems for collection of Logs, Metrics, and Tracks - The ability to ingest telemetry collected from hosted or SaaS-based applications		Y	
Instrumentation-based Aggregation of Logs, Metrics and Tracks at the services and issues level		Y	
Real-Time Monitoring of data pipelines, data transformations, and data interactions to ensure data quality and consistency		Y	
Contextual, proactive, and dynamic Anomaly Detection which involves detecting "unknown unknowns" anomalies, exceptions events, and unexpected patterns in data that may indicate data quality issues or errors		Y	
Adding Domain Context and Metadata in business activity monitoring, analysis of KPIs, Endpoint Monitoring, user journey/experience and its impact on business outcomes		Y	
Root-Cause Analysis		Y	
Performance Tracking and tracking latency of data pipelines and processes to ensure that data is delivered within acceptable timeframes		Y	
ML/AIOps features such as identification of recommendations for resolving health, self-healing, capacity forecasting, and performance anomalies using advanced analytics		Y	
Integration with DevOps – toolchains to facilitate continuous or progressive application delivery		Y	
Performance testing and integration with load testing tools		Y	
Data Policies and Standards to establish data policies, standards, and guidelines to ensure that data is managed consistently and meets quality requirements			Y
Data Stewardship – assigning roles and responsibilities for data stewardship, ensuring that data owners and custodians are accountable for data quality and integrity			Y
Data Compliance – ensuring data management practices adhere to regulatory and compliance requirements, safeguarding sensitive and personal data			Y
Data Lifecycle Management from data creation and ingestion to archival and disposal			Y
Data Access and Security – managing data access controls, permissions, and encryption to protect data from unauthorized access and breaches			Y

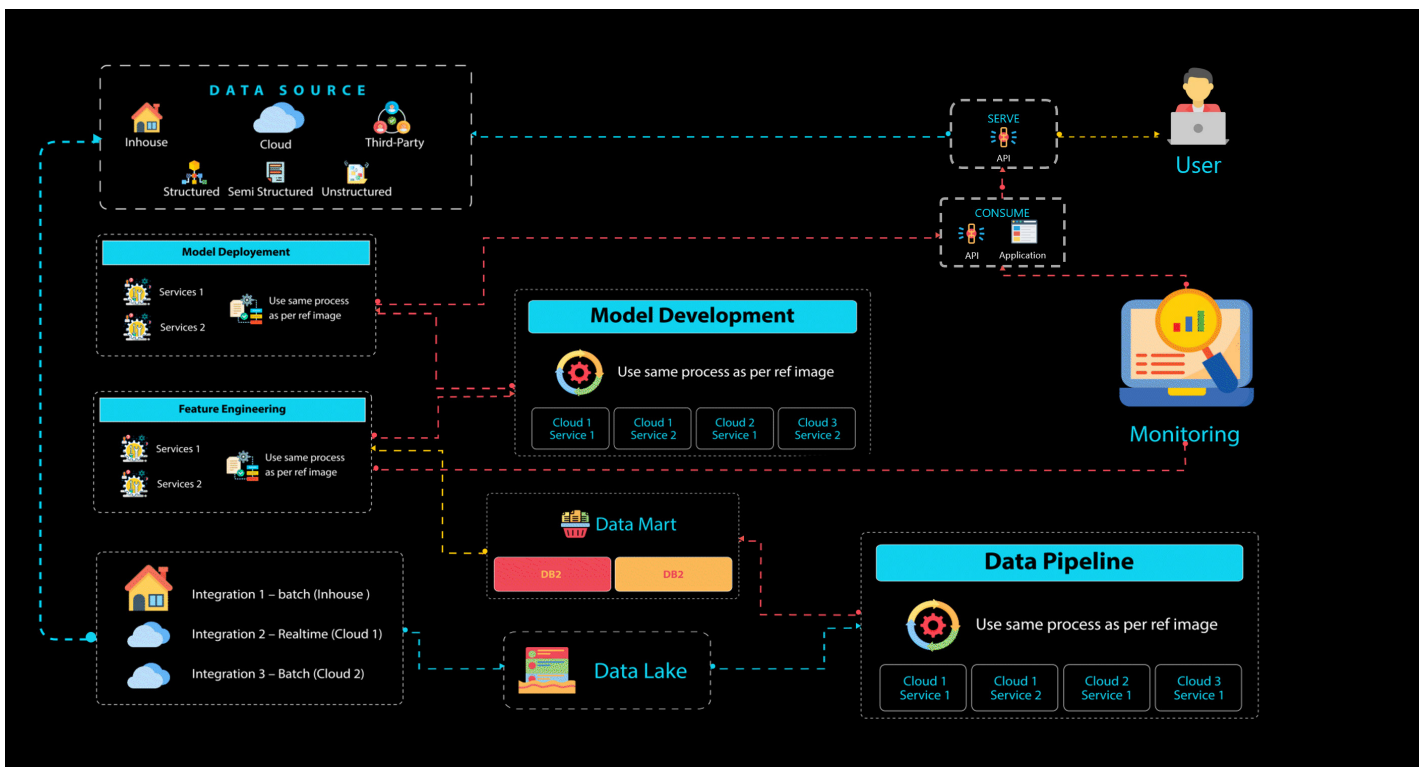
Why Data Observability

In the evolving landscape of enterprise operations, where data informs crucial decisions and underpins analytics, ensuring data accuracy becomes paramount. Discrepancies or outdated records have the potential to bewilder stakeholders and lead to detrimental actions. Consider instances like a machine learning model wrongly flagging legitimate transactions as fraudulent or a sales manager engaging with a customer unaware of their recent support desk complaint. Such scenarios underscore that erroneous or incomplete data poses a risk rather than an asset.

To transform data into an valuable asset, enterprises require the adoption of Data Observability.



Data Observability assumes greater significance in the context of contemporary D&A platforms with constant evolution, high dynamism, complexity, and distribution. They amalgamate diverse architectural patterns, service tiers, and an array of tools and technologies, as depicted in the accompanying graphics. Managing these elements necessitates a comprehensive and integrated approach to observe, proactively alert, pinpoint root causes of failures, and provide governance-oriented recommendations. This holistic endeavor is essential for achieving overarching business objectives in this modern landscape. Top of Form



Diverse Platform Architecture

As organizations start adopting data observability, they unlock multiple advantages that extend beyond system health, directly impacting their core business operations.

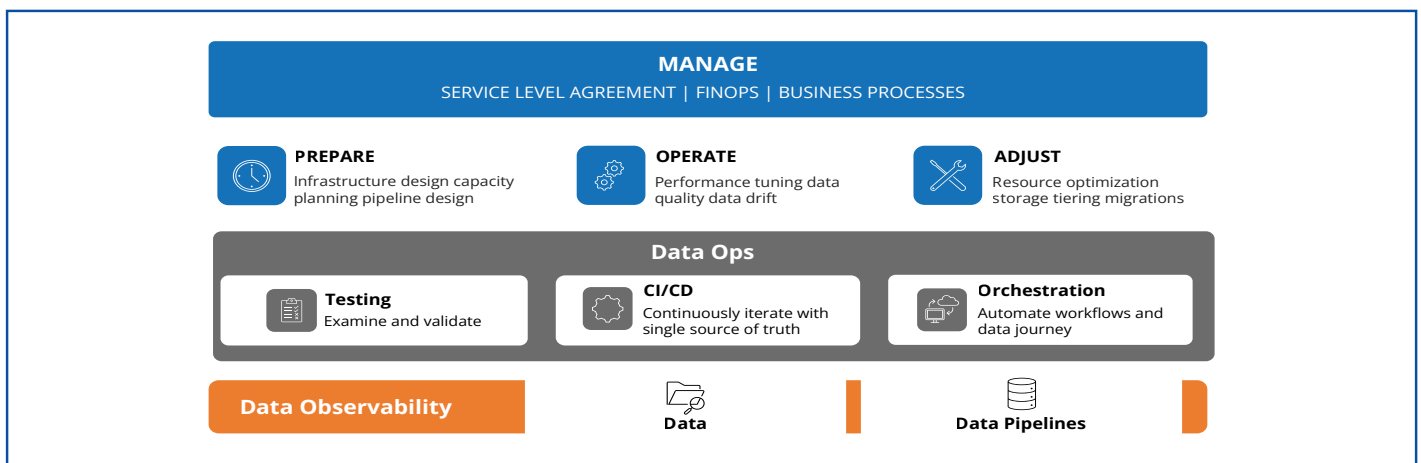
Here are compelling reasons why organizations should embrace Observability:

- **Improved System and Data Reliability:** Observability acts as a proactive guardian, enhancing the reliability of both data and systems. It identifies and addresses issues before they escalate into outages.
- **Enhanced User Experiences:** By pinpointing and remedying performance bottlenecks, Observability elevates user experiences. It instills trust by meticulously managing data quality through rigorous business validation.
- **Empowered Business Decision-Making:** Observability empowers organizations to construct domain-centric data products. It enables self-service analytics, fosters knowledge sharing, facilitates data dissemination, and boosts data value and adoption, thereby facilitating superior business decision-making.
- **Accelerated Time to Market:** Observability expedites the production of data products and streamlines data access through robust testing, data contracts, and efficient deployment processes.
- **Reduced Costs of Ownership:** By preventing outages and swiftly addressing data issues, Observability curtails costs. It establishes clear accountability, optimizes performance, and streamlines infrastructure and data operations.
- **Informed Technical Decision-Making:** The collection and analysis of system data enable a deeper understanding of system functionality, facilitating informed decisions on optimization and improvement.
- **Proactive Operations:** Observability obviates the need for extensive debugging in deployment environments, reducing the necessity for code changes. It promotes agile development and enables teams to save time through data-driven transformation and collaborative practices, such as xOps.
- **Enhanced Security:** Observability plays a pivotal role in bolstering security by identifying and addressing threats and affording better control over cloud data.

In summary, embracing Observability not only enhances the robustness of systems and data but also directly influences user experiences, business decision-making, time-to-market, cost efficiency, technical proficiency, proactivity, and security measures. It emerges as an indispensable asset in the modern organizational landscape, with varied use cases.

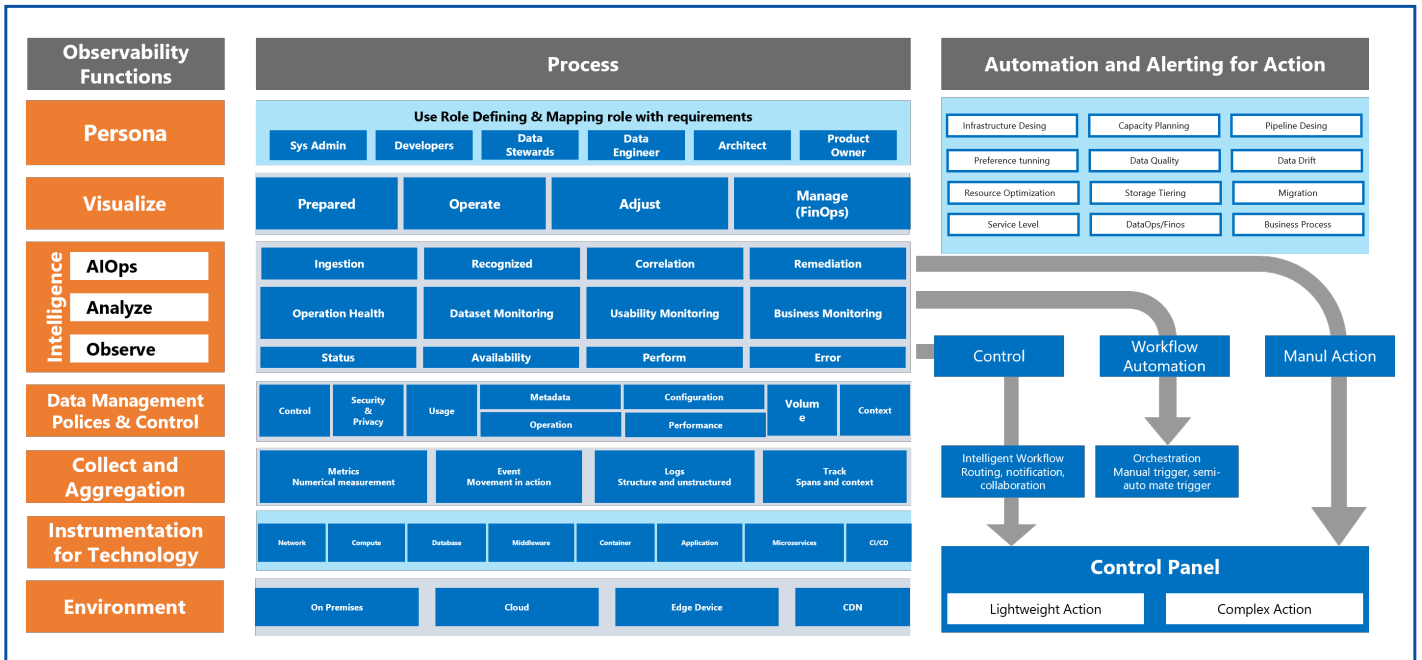
Use Case of Data Observability?

To achieve the above benefits, Observability enables the below-mentioned use cases.



Use Case of Data Observability

- **Prepare:** Data observability helps data teams prepare the data environment. The primary use cases for this category are designing their infrastructure, planning the capacity of infrastructure resources, and creating the pipelines that deliver data for consumption.
- **Operate:** Data observability also helps data teams operate their environment. The use cases for this category include studying and tuning pipeline performance, finding and fixing data quality issues to ensure adequate data quality management, and identifying data drift that affects machine learning (ML) models.
- **Adjust:** Next, data observability helps data teams adjust the data environment. This category includes use cases such as resource optimization, storage tiering, and migrations
- **Fund:** Finally, data observability helps business and IT leaders fund analytics projects and applications from a business perspective. This category of use cases focuses on the Financial Operations (FinOps) use case.



Observability Frameworks

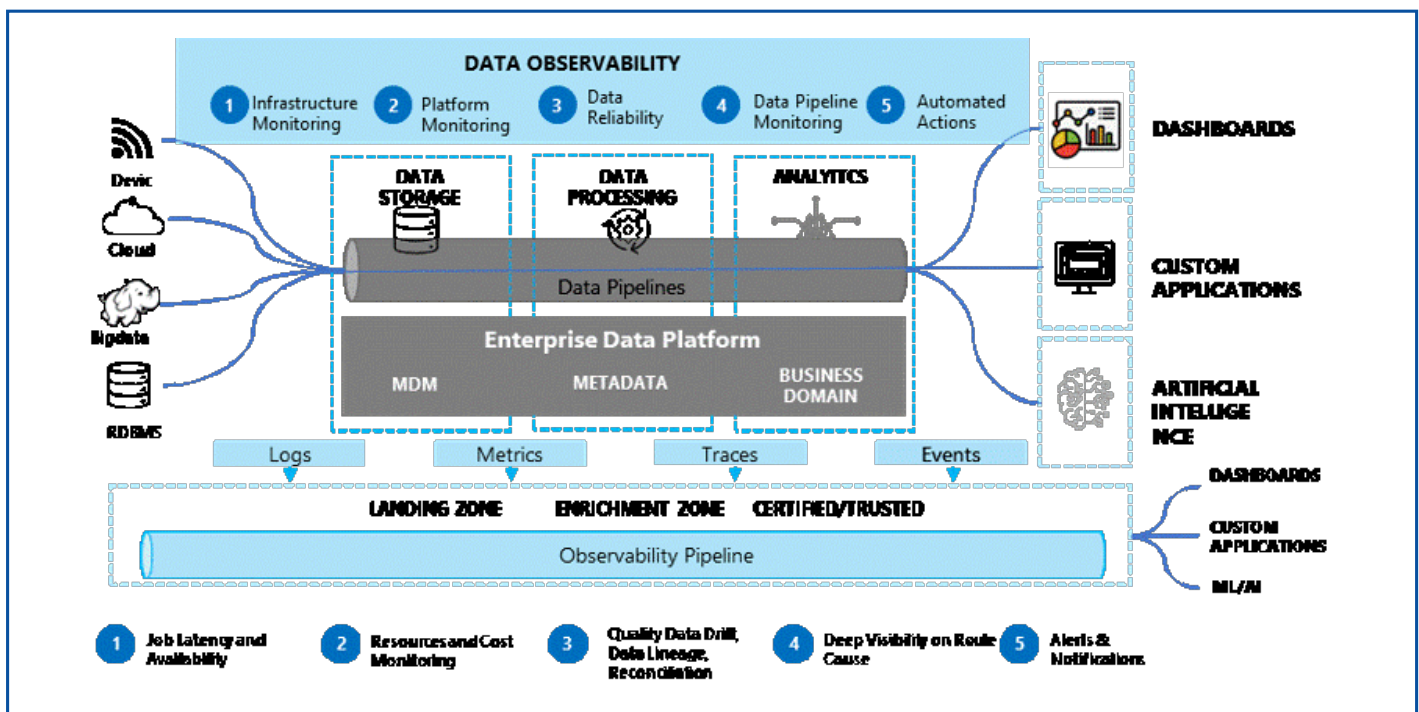
The key components of Data Observability are:

Data Observability comprises several core components, each playing a vital role in comprehensively understanding a system's internal dynamics and behaviors. These components are:

- **Data Collection:** Gathering essential data about the system's lineage, state, and behaviors is fundamental. This involves collecting diverse data types, including:
 - **Metrics:** Numeric measurements reflecting the system's condition, such as request rates or response times.
 - **Logs:** Textual records capturing system events like errors or exceptions.
 - **Traces:** Sequences of data points tracking the path of a request through the system.
 - **Events:** Data points representing discrete occurrences like user logins or database queries.
- **Instrumentation:** This entails incorporating code or leveraging third-party tools to facilitate data collection within the system.
- **Common Data Schema:** Storing collected data in an accessible and analyzable format is essential. This involves adopting a well-defined information architecture that prioritizes data consistency and quality.
- **Metadata:** Additional contextual information associated with the data, such as collection time, source, or the user responsible for data acquisition.

- **Data Contextualization:** Beyond mere data collection, observability centers on understanding the context surrounding the data. This encompasses discerning relationships between various data points, understanding data meaning, and recognizing its significance.
- **Data Correlation and Analysis:** Analyzing the collected data to uncover patterns and trends is crucial for grasping the system's state and behavior. This involves employing diverse tools and techniques, including machine learning, statistical analysis, and natural language processing.
- **Visualization:** Effectively presenting analyzed data is imperative. This is achieved through various tools and technologies such as dashboards, charts, and graphs, making complex information readily understandable.
- **Automation and Alerting:** Setting up automated alert systems serves as a proactive measure to promptly identify and address issues. These alerts aid in swift troubleshooting, enhancing overall system reliability.

The core components of Data Observability form a comprehensive framework for gathering, contextualizing, analyzing, and presenting system data. This approach empowers organizations to gain valuable insights into system behaviors, enabling informed decision-making and the swift resolution of issues.



However, Data Observability implementation can be challenging

The pursuit of effective Data Observability encounters several challenges, encompassing:

- **Data Explosion and Scalability:** The escalating volumes, varieties, and velocities of data stemming from diverse sources and digital interactions can overwhelm data pipelines. Collecting, storing, and promptly analyzing Observability data can pose significant logistical hurdles, making it challenging for organizations to justify the requisite investments.
- **Standardization of Data Sources:** Large organizations contend with a multitude of data sources, often adhering to disparate standards. While observability tools aim to standardize telemetry data and logging guidelines for effective correlation, normalizing data frequently demands manual effort.
- **Noise:** The prolific generation of logs and metrics from pipelines and infrastructure creates a cacophony of data, drowning out critical signals pertaining to data health.

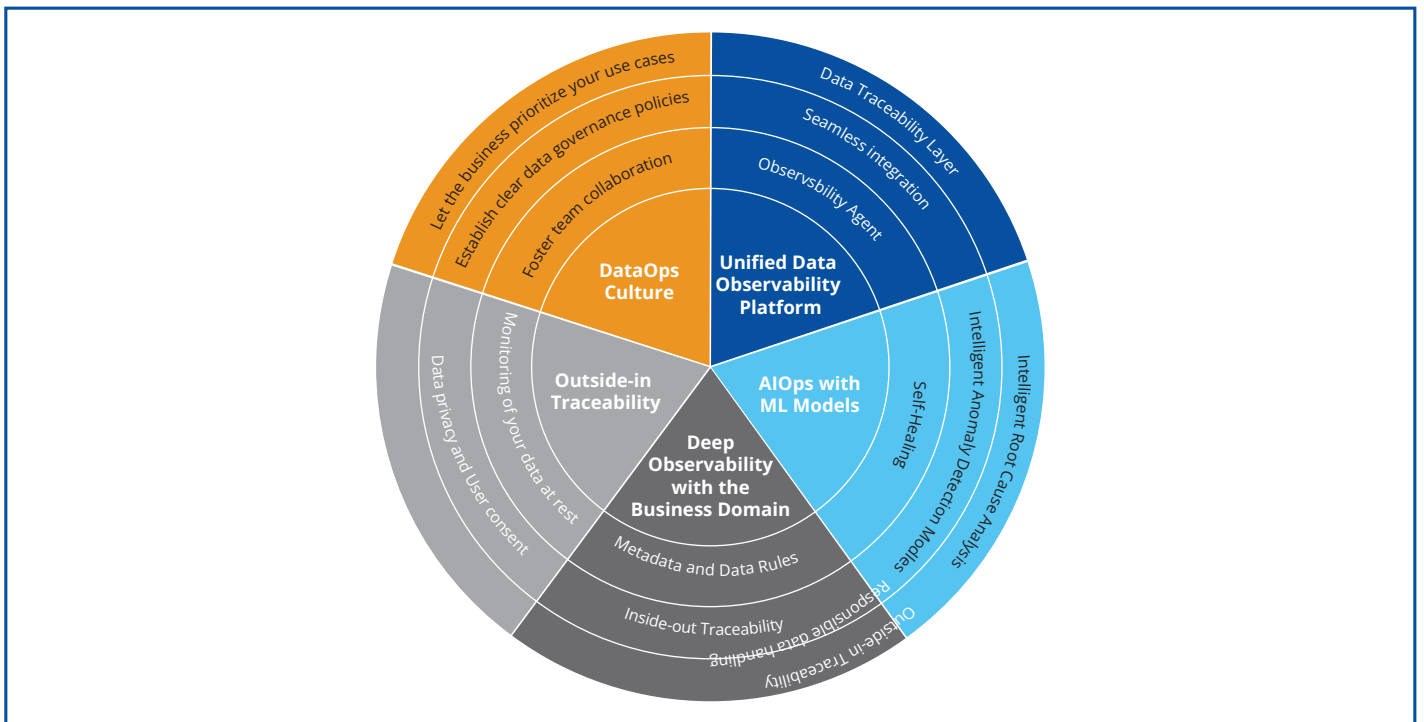
- **Isolated Signals:** Data teams and engineers, particularly, grapple with the task of connecting isolated events to comprehend sequential occurrences effectively.
- **Fragmented Tools:** Many tools offer fragmented views, with some focusing on operational aspects rather than analytical workloads. This fragmentation can hinder comprehensive Observability.
- **Integration with the Entire Data Ecosystem:** Even the most proficient observability tools may fall short without insight into the entirety of the data pipeline, including all associated software, servers, databases, and applications. Eliminating data silos and integrating as many systems as possible into the data observability software is imperative. However, organizations may encounter resistance when seeking buy-in to incorporate all systems and tools.

Defining the NorthStar for implementing Data Observability for your Organization.

These challenges can be mitigated using the proper tools and techniques and having a well-defined Observability strategy.

Here are some specific tips for mitigating Observability challenges:

North Star for Implementing Observability

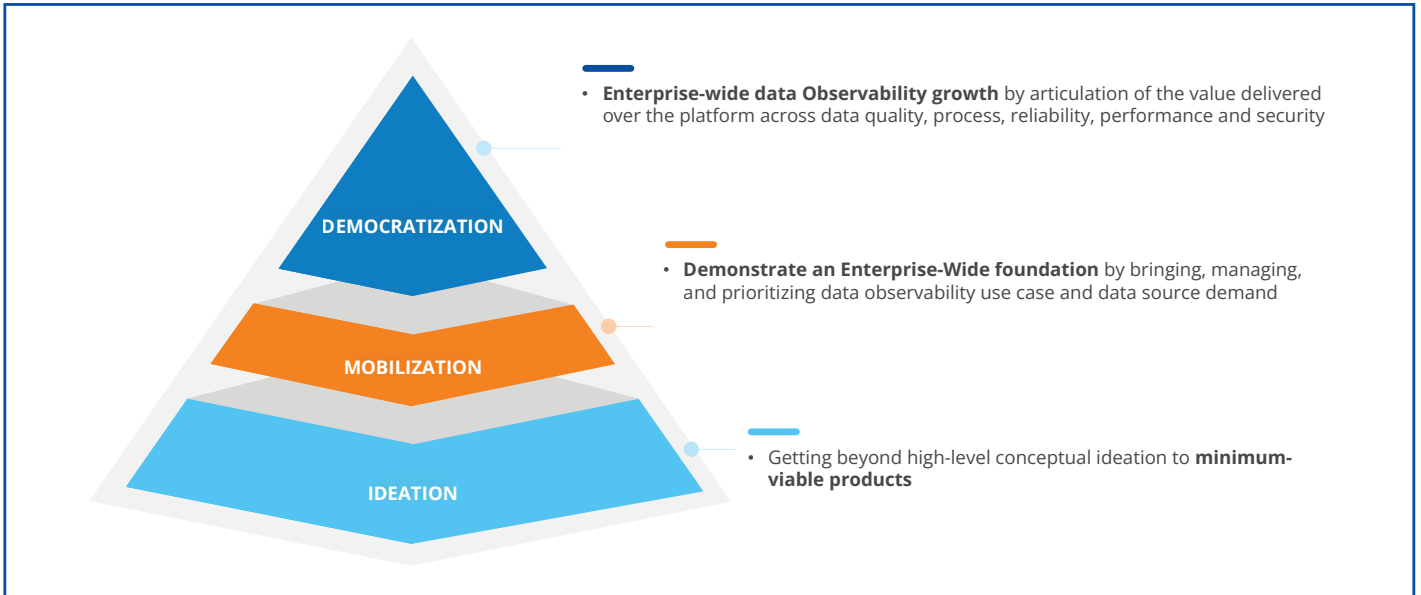


- **DataOps Culture:** Establishing a pervasive DataOps Culture is a prerequisite for realizing the full potential of Observability. It hinges on universal recognition of the importance of operational efficiency. Once this mindset permeates, organizational leadership can strategically steer towards a standardized data platform and DataOps system. Achieving the right equilibrium between delivery cost and value is paramount. This approach cultivates end-to-end ownership and encourages open communication among teams. Additionally, it underscores the significance of engaging business stakeholders in the prioritization of use cases, ensuring alignment with broader organizational objectives.

- **Unified Data Observability Platform:** A centralized Observability platform serves as a linchpin in the quest for comprehensive insights. It streamlines the collection, storage, and analysis of data drawn from multiple systems, offering a panoramic view of the entire ecosystem and expediting root cause identification. Notable attributes of an effective unified platform encompass:
 - **Observability Agent:** Facilitating swift and seamless integration with existing technology stacks.
 - **Seamless Integration:** Seamless integration with enterprise data governance tools, featuring a standardized data schema.
 - **Data Traceability Layer:** Enriching contextual understanding and proactively addressing data reliability concerns.
- **AIOps with ML Models:** Harnessing the capabilities of machine learning models within data observability tools introduces automation and precision to the problem identification process. This approach is instrumental in streamlining operations by reducing false positives. Key facets include:
 - **Self-Healing:** Automation of issue remediation through the creation of incident tickets, their routing to the relevant teams, and execution of scripts to address issues across cloud infrastructure and data processes.
 - **Intelligent Anomaly Detection Models:** Leveraging machine learning models to scrutinize data trends, seasonality, and patterns, detecting anomalies and minimizing the Mean Time to Detect.
 - **Intelligent Root Cause Analysis:** Employing machine learning-driven event correlation and causality analysis to pinpoint probable root causes. This contextualizes issues within a single framework, enhancing the Mean Time to Resolve.
- **Deep Observability with the Business Domain:** Incorporating domain context into Observability infuses it with real-world relevance, extending its utility to business stakeholders, partners, and regulatory entities. Key facets encompass:
 - **Metadata and Data Rules:** Facilitating the implementation of customized validation rules, ensuring alignment with unique business logic and requirements. This can be achieved through Metadata, Glossary, and Data Enrichment Rules, fostering deep data observability tailored to specific business needs.
 - **Inside-out Traceability:** Empowering users to understand how Data Platform errors impact Business KPIs, enabling informed decision-making and compliance assessments.
 - **Outside-in Traceability:** Providing insights into how operational platform failures or transactional issues, such as downtime in e-commerce, CRM, or ERP platforms, affect data platform performance.
- **Ethical and Privacy Considerations:** Incorporating Observability while upholding ethical and privacy standards is non-negotiable. This necessitates a meticulous approach to data handling and security to meet the most stringent requirements. Key considerations include:
 - **Monitoring at Rest:** Enabling data monitoring without extracting it from its current storage location, ensuring optimal security, performance, scalability, and cost-effectiveness.
 - **Data Privacy and User Consent:** Instituting protocols to inform users about data collection and its use, respecting their privacy rights.
 - **Responsible Data Handling:** Adhering to responsible and secure data management practices, ensuring data integrity and compliance with regulatory frameworks.

By effectively addressing these multifaceted challenges, organizations can bolster the efficacy of their Observability endeavors, culminating in enhanced system reliability, performance, security, and stakeholder experiences.

North Star for Implementing Observability



- **Ideation**

- Begin by defining use cases and aligning success metrics with business priorities. Key use cases encompass Data Catalog, DPDQ (Data Profiling and Data Quality), Data Redundancy, and Data Use Adoption, each with its automated processes for data management.
- Leverage the success metrics to define data management frameworks for each charter of Data Observability.
- Use features from the above framework for the evaluation of tools, technology, people, and processes.
- Getting beyond high-level concepts to minimum-viable products and evaluating it.

- **Mobilization**

- Create a strong foundation for Data observability success and rapid experimentation to empower every use case with easy-to-configure and scalable.
- Demonstrate an Enterprise-Wide foundation by bringing, managing, and prioritizing data observability use case and data source demand.
- Enabling DataOps Culture, Unified Data Observability Platform, AIOps with ML Models, Deep Observability with the Business Domain, Ethical and Privacy Considerations.

- **Democratization**

- Enterprise-wide data Observability growth by articulation of the value delivered over the platform across data quality, process, reliability, performance and security
- Foster scaling and collaboration by enabling Governance, COE Factory, Knowledgebase, and Community
- Set the benchmarks and innovation by collaborating with partners and technology vendors.

Course5 Case Studies: Driving Impact through Data Observability

- Availability of Data Lineage part of data observability enabled migration and operation pathway by defining rationalization, optimization, and simplification strategy, driving 75% cost to serve benefits.
- Data Observability for an eCommerce website helped improve Product Data Quality by 45% and Customer Experience (NPS) by 21%.
- Data Observability reconciliation capability uplift worth £1.7M in potential sales identified based on missing SKUs on physical platforms by Reconcile AI with online and offline MDM.

Future Directions and Trends

The future of data observability is being shaped by several prominent trends and innovations, each poised to enhance its power and effectiveness.

- **AI-driven Insights:** Artificial intelligence (AI) holds the potential to analyze observability data, uncovering intricate patterns and trends that may elude manual detection. This promises faster and more accurate problem identification and troubleshooting.
- **Greater AI and ML Integration:** Observability tools are increasingly integrating artificial intelligence and machine learning capabilities, enabling advanced data analysis and anomaly detection. This accelerates issue identification and system optimization.
- **Context-aware Observability:** Incorporating data context into analysis—considering other data, the environment, and user objectives—enhances the accuracy of generated insights.
- **Observability in Edge Computing:** Deploying computing resources closer to end users in edge computing can enhance application performance and responsiveness. Observability plays a pivotal role in monitoring and managing these systems.
- **Security and Privacy Focus:** As data security and privacy gain paramount importance, observability tools are incorporating features like data masking to safeguard sensitive information.
- **Open-source Observability Tools:** Organizations are increasingly adopting open-source observability tools for their agility and customizability. These tools can be extended rapidly, making them attractive for monitoring diverse data sources.

Conclusion - The Call to Action

In an era defined by interconnectedness and digital transformation, traditional methods of system monitoring, and maintenance must evolve to meet the demands of modern applications. Achieving trustworthy insights in real-time is imperative to ensure reliability, performance, and exceptional user experiences. Observability, originally rooted in DevOps and system administration, has transcended boundaries, reshaping how we perceive, manage, and glean insights from complex systems. By providing the means to gain a comprehensive understanding of a system's inner workings, applications, and processes, Observability has become the cornerstone for informed decision-making and proactive issue resolution. Embracing this paradigm shift is the call to action for organizations seeking to thrive in the ever-evolving landscape of digital systems and data-driven decision-making.

About Course5 Intelligence

Course5 Intelligence Limited ("Course5") focuses on helping organizations drive digital transformation using artificial intelligence ("AI"), advanced analytics and insights. Course5's AI-driven products and solutions and IP-led solutions are supported by industry-specific domain experience and the latest technologies and aim at enabling organizations to solve complex issues relating to their customers, markets and supply chain at speed and scale. Course5 combines a multi-disciplinary approach to data integration across structured and unstructured data sources to help businesses grow through informed decision-making.

Course5 caters to some of the world's largest enterprises, including many Fortune 500 companies. The company's clients span Technology, Media and Telecom (TMT), Pharma & Lifesciences, CPG, Retail, and other sectors. Course5 Intelligence has been recognized by leading industry analysts like Gartner and Forrester for its Analytics and AI capabilities and proprietary AI-based platforms.

Contact Us



Visit : www.course5i.com