LTIMindtree

WHITEPAPER

Data Observability

Faster production of trustworthy financial data

Executive Summary

Organizations are setting up financial data hubs, as part of their financial transformation, which works as a single source of truth. The data is pulled from multiple business units such as finance, Human Capital Management (HCM), procurement, and sales and marketing. This data is used across business outcomes, including top-line sales, Profit and Loss (P&L), manpower cost, direct and indirect tax, Goods and Service Tax (GST), and corporate tax. Similarly, it's extremely critical to produce trustworthy data for annual, quarterly, monthly, and weekly reporting to the various stakeholders, not limited to leadership or internal team, but the external bodies such as regulatory organizations. Any error in the report may jeopardize the company's reputation, attract attention from regulatory authorities, and lead to monetary loss.

Data observability can reduce these risks in financial data by continuously observing it and detecting any anomaly, even before reconciliation with data in transactional systems, ledgers, or journals. This would help reduce the close cycles and minimize the reconciliation efforts. Data observability involves real-time monitoring, automated alerts, and data validation to ensure data quality and accuracy. Its impact is far-reaching. With correct and reliable data, organizations can make informed decisions and improve financial & non-financial reporting.

In this whitepaper, we'll explain what are some of the challenges that organizations face when implementing data observability and the framework you can use to assess and implement data observability.

What is data observability?

Data observability involves enabling organizations to gain a comprehensive view of their data ecosystem, including intricate data connections like data pipelines, infrastructure, and applications. The primary goal is to swiftly detect, manage, prevent, escalate, and resolve data disruptions within predefined service level agreements (SLAs). Data observability relies on continuous collection, consolidation, and analysis of signals across multiple layers to achieve these objectives. It also informs and recommends improved designs for enhanced performance and aligns data governance with business objectives.

By the year 2026, it is anticipated that 70% of organizations effectively leveraging observability will experience reduced decision-making latency, thus gaining a competitive edge in their business or IT operations. The essence of observability lies in using surface-level data to gain insights into the internal data landscape. Observable data holds significance because it is rooted in confirmed stakeholder actions rather than mere intentions, obligations, or promises, rendering it a solid foundation for evidence-based decision-making.

Achieving data observability

Data observability can be achieved through a combination of processes, practices, and technology. To make data observability useful, it needs to include activities such as monitoring, alerting, tracking, comparisons, and analysis. To excel at data observability, one needs to collect as much information about what's happening within the data platform as possible. You can attain that information through data quality monitoring, regularly checking data against pre-defined data quality rules, and presenting results and trends. To implement data observability correctly, you should start with an observability strategy and framework to prepare your team for its impact on their workflows.

For e.g., if you need the right data at the right time for the Profit and Loss (P&L) reporting, it is necessary to monitor the data refresh between the General Ledger (GL) systems and the finance hub. It ensures that daily P&L is published in the morning, and if there is a delay in any of the translation or refresh processes, then communication is enabled with the stakeholders to avoid surprises.

Challenges on the way to achieve data observability maturity

Achieving data observability maturity is a complex endeavor for organizations, as it involves overcoming various challenges throughout the process. Here are some key challenges they may face:

1. **Siloed data and systems:** Organizations often struggle with data scattered across different systems and departments, resulting in fragmented data landscapes. Achieving data observability requires breaking down data silos and integrating disparate data sources to gain a holistic view of the organization's data. For the finance team, the data would reside between multiple systems, such as GL, HCM, CRM, etc. Below is a pictorial representation of how a single piece of information, staff cost, is treated differently by different business units.

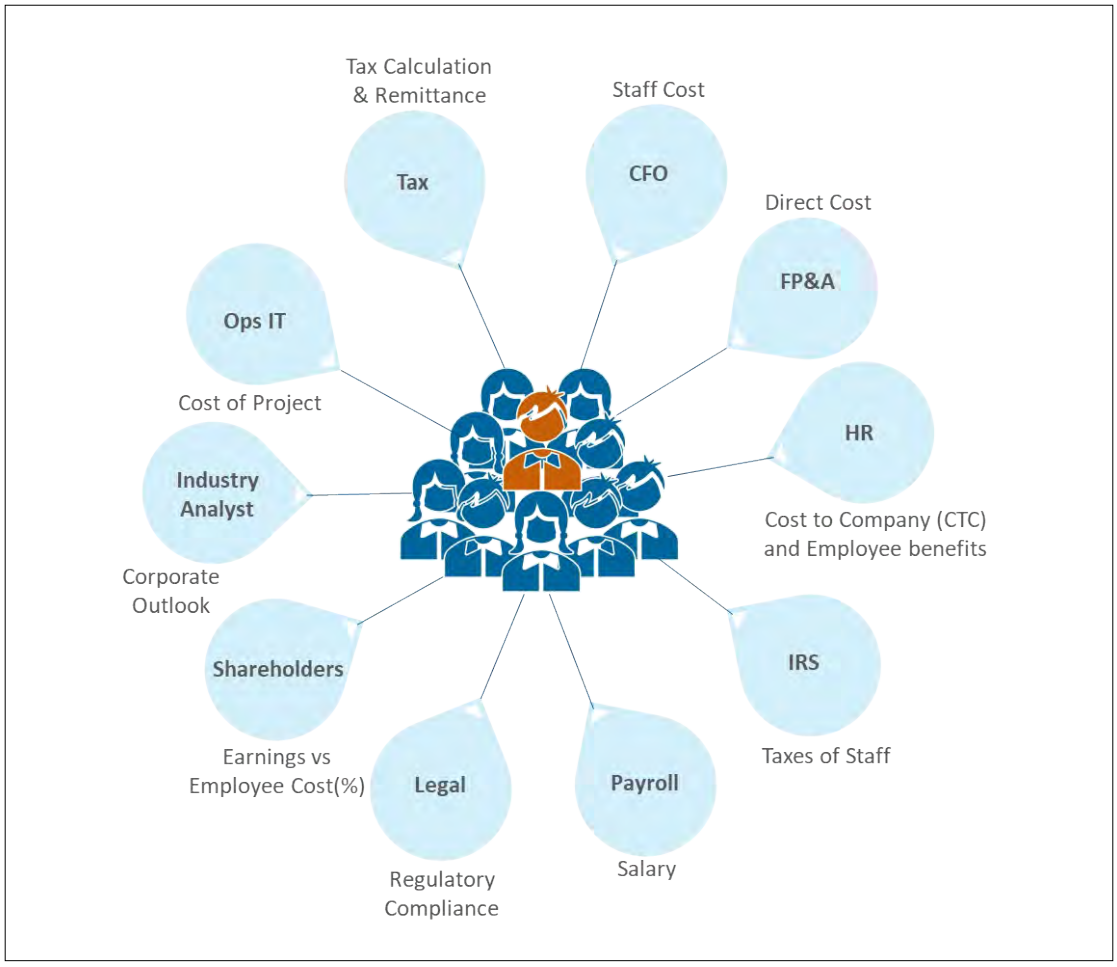

Figure 1: How staff cost is treated differently by different business units

2. **Lack of data quality and consistency:** Poor data quality and inconsistencies pose significant challenges to data observability. Incomplete, inaccurate, or outdated data can lead to flawed insights and decision-making. Organizations must invest in robust master data, data cleansing, standardization, and validation processes to ensure data integrity and consistency.

3. **Complex data ecosystems:** Modern data ecosystems encompass diverse data types, formats, and technologies, including cloud platforms, streaming data, APIs, and external data sources. Managing and monitoring these complex environments poses challenges in terms of data integration, interoperability, and compatibility, making it difficult to establish comprehensive data observability.

4. **Scaling data observability:** As data volumes continue to grow exponentially, organizations face scalability challenges in monitoring and analyzing large datasets in real-time. Traditional tools and approaches may become inadequate, requiring the adoption of scalable technologies and automated processes to keep pace with the increasing data demands. As the finance teams need to collaborate with other BU, the quantum of operational data is high, so it is necessary to have a model scalable solution.

5. **Limited visibility and contextual understanding:** Understanding the context and lineage of data is crucial for effective observability. However, organizations often lack visibility into data origins, transformations, and dependencies. Establishing robust data governance frameworks, metadata management practices, and data lineage tracking mechanisms is essential to overcome this challenge. While the finance team is primarily driven by the GL(General Ledger) data, it is also necessary to bring the context of transactional data, such invoices or sales orders to ensure synch between what is being transacted vs. booked.

6. **Organizational culture and collaboration:** Cultivating a data-driven culture and promoting collaboration between IT and business functions can be a significant hurdle. Siloed mindsets, resistance to change, and a lack of awareness about the benefits of data observability can impede progress. Organizations must foster a culture that values data transparency, collaboration, and continuous improvement. Finance being business-driven, it is necessary to have the right collaboration between business and IT to get the right outcomes with data.

7. **Resource constraints:** Implementing data observability initiatives requires substantial investments in technology infrastructure, skilled personnel, and ongoing maintenance. Limited resources, budgetary constraints, and talent shortages can hinder the organization's data observability maturity.

8. **Overcoming these challenges requires a robust framework and comprehensive approach that combines technology,** processes, and cultural shifts. Organizations should prioritize data governance, invest in scalable technologies, establish data quality frameworks, foster collaboration, and align data observability initiatives with strategic business goals. By addressing these challenges, they can unlock the full data potential, ensure trustworthiness, and leverage actionable insights for informed decision-making.

Solution to the challenges

To address these data observability needs effectively, organizations should adopt a multifaceted approach. The first steps to this would be answering the question of "do we get right data at right time?". This requires strong collaboration with business and IT, who can validate multiple systems and their data usages. While the business can define the process point connects on how a specific data in one of the system is being leveraged for other departments system (eg: the HR department's definition of salary and how payroll/finance team access it and build their dataset and then pass it to taxation team). By enabling the process points integration, it is easier for the observability process to identify the failure of HCM system, how it impacts the payroll team during the month end processes.

While the IT team would work on breaking down data silos through data integration and centralization, establishing rigorous data quality frameworks and governance, implementing data catalogs and metadata management for comprehensive understanding, investing in scalable observability tools to handle data growth, creating contextual visualization dashboards, fostering a data-centric culture, encouraging cross-team collaboration, and optimizing resource allocation by prioritizing critical data observability needs

These initiatives collectively ensure improved data quality, operational efficiency, enhanced collaboration, and prudent resource utilization, empowering organizations to make data-driven decisions in the modern data landscape.

These multifaceted solutions offer a wide array of advantages:

- Enhanced data quality and precision analytics that are essential for better decision-making.
- Improved understanding of complex data ecosystems leads to streamlined operations and prioritization of systems based on the period of month.
- Contextual visualization empowers informed outcomes. Fostering a data-centric culture and encouraging cross-team collaboration sparks innovation and bolsters competitive advantage.
- Last but not the least, an integrated data observability solution empowers business users to work optimally on specific datasets for a given period rather working on all the datasets. This would translate to tangible business benefits such as improved efficiency, ehanced productivity, and building new strategical opportunities using data than working on operational activities.

Data observability maturity framework

- Here's how you can evaluate the data observability maturity of an organization.
- Assess data-driven culture: Assess the implementation of data-driven culture by evaluating whether they encourage the use of data for decision-making through reports, dashboards, and analysis and promote data literacy among employees.
- Review data governance framework: Review the implementation of a data governance framework to ensure that data is managed consistently across the organization. Check whether data ownership, data lineage, and data usage policies are defined.
- Data observability metrics: Evaluate the implementation of data observability metrics, including data completeness, timeliness, accuracy, consistency, integrity, availability, and security. Check whether these metrics have been defined and implemented consistently across all data sources.
- Identification of key data sources: Confirm if critical data sources within the organization, such as transaction, customer, and operational data, are identified for observability.
- Evaluate data quality checks: Evaluate the implementation of automated data quality checks to ensure compliance with the defined observability metrics. Check whether data validation rules, data profiling, and data cleansing processes are implemented consistently.
- Analyse data monitoring: Check if data is continuously monitored to ensure that it remains within the defined observability metrics. Check whether alerts and notifications are set up to inform stakeholders of any deviations from the expected values.
- Evaluate data analytics tools: Evaluate the investment in data analytics tools for analysis of the observed data. Check whether data visualization, data modeling, and data analysis platforms are implemented.

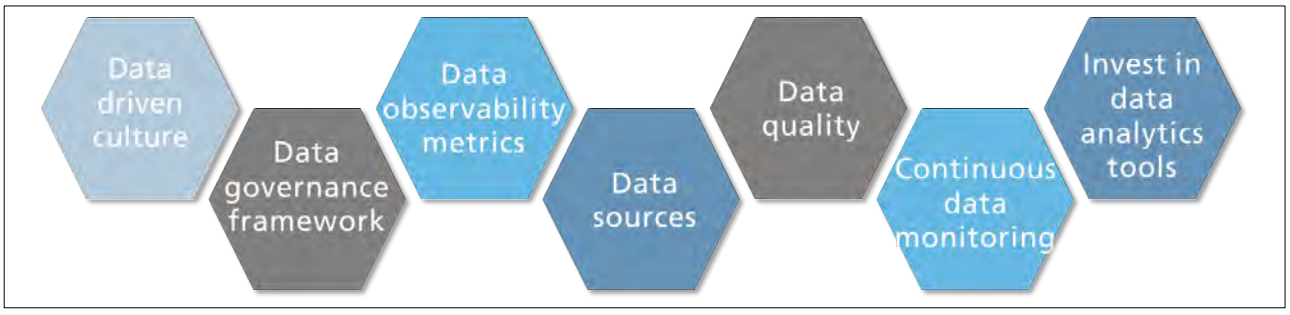

Figure 2: Elements of data observability

Based on the evaluation of these steps, you can determine the maturity level of the company's data observability implementation. The evaluation will detect the current state of data observability maturity, find areas for improvement, and guide the development of a strategy and roadmap for the company's journey toward data observability maturity.

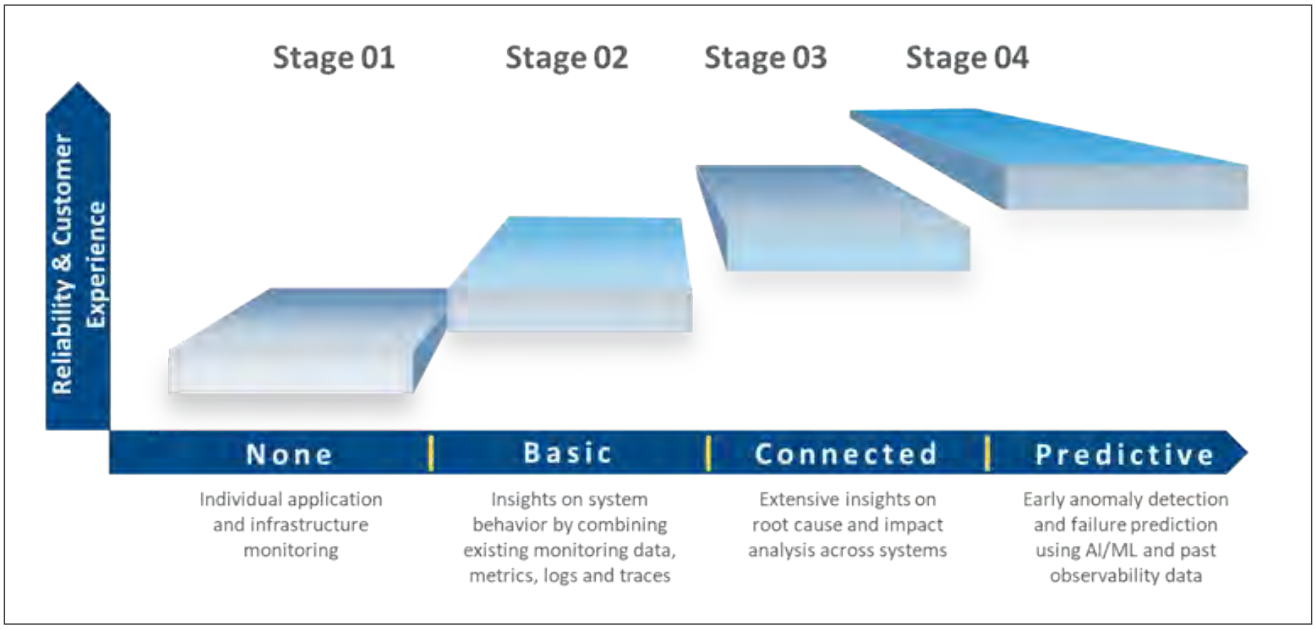

Figure 3: Observability maturity vs. IT reliability and customer experience

Here is the checklist for measuring the level of maturity data observability

Maturity stages	Evaluation criteria
Stage 1 - None	• Is health check available for individual IT systems only? • Are event-based triggers available for IT systems? • Do you get notifications and alerts in case of errors and warnings? • Which channels are used to communicate the alerts?

Maturity stages	Evaluation criteria
Stage 2 - Basic	• Is there a real-time dashboard for insights into system observability? • What are the components of the observability dashboard – logs, traces, and metrics? • Do you connect historical log data with the current monitoring systems? • Is there any report/dashboard for root cause analysis and impact analysis? • Which tool are you using for real-time observability reporting? • Do you observe data 24/7? • Have you identified and observed all the data sources?
Stage 3 - Connected	• Is there a process to perform comprehensive causal analysis spread across systems used by the organization? • Is there a system that provides extensive, correlated information on what went wrong, when and why the issue occurred, and what other areas were affected? • Does this information help to overcome the breakdown faster? • Do you check data quality parameters within data observability? • Is there a list of well-defined and documented metrics used for data observability? • How does a data governance team monitor data observability activity? • Are you storing all the observability incidents for further reference?

Maturity stages	Evaluation criteria
Stage 4 - Predictive	• Are you using AI/ML on observability data to detect patterns? • Is the system capable of detecting anomalies during the start of the incident? • Is automated alert and notification available for data observability? • Can your system predict a potential incident and notify accordingly?

Data observability metrics

Data timeliness: This measures how fresh the data is. Timely data is important for making informed decisions in real time.

Data accuracy: This measures how correct the data is. Accurate data is essential for making informed decisions and preventing errors.

Data consistency: This measures how consistent the data is across different sources. Inconsistent data can lead to errors and inconsistencies in analysis.

Data integrity: This measures the overall quality of the data. Data with high integrity is reliable and trustworthy.

Data availability: This measures the availability of the data. Availability is critical for ensuring that data is accessible when needed.

Data security: This measures the security of the data. Security is essential for protecting sensitive information and preventing data breaches.

These metrics are used to assess the overall observability of data within a system. By monitoring and measuring these metrics, organizations can ensure that their data is complete, correct, and available for analysis when needed. This helps to improve the overall quality and reliability of insights derived from the data.

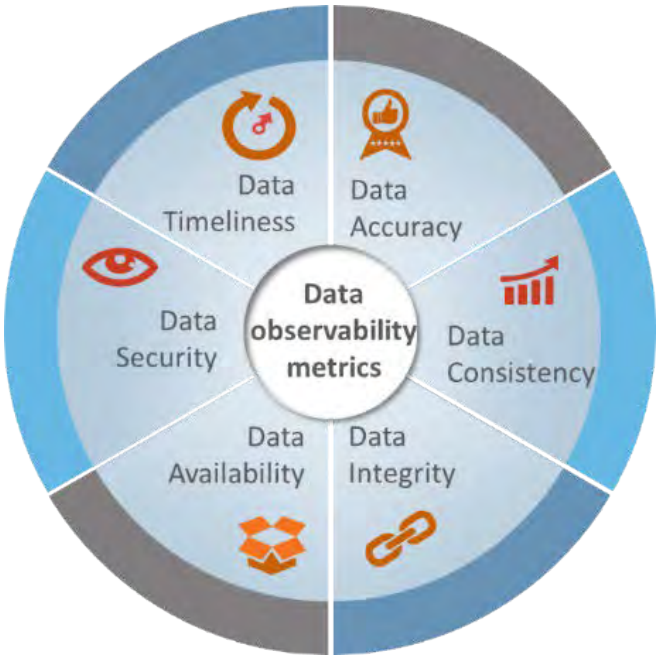

Figure 4: Data observability metrics

Metric definitions

Here are the logic of calculating the KPIs to measure the matrics related to data quality, freshness, completeness, and accuracy.

Metric	Description and data observability score calculation logic
Dataset	**Description:** Average of all the scores across all the attributes and Data observability checks
	Logic: Row wise score + Column level score for each checks and average of total score

Metric	Description and data observability score calculation logic
Uniqueness	**Description:** Measure of the uniqueness of the dataset as a whole and at column level if unique columns defined. Columns having unique values after removing NULLs are considered for this check
	Logic: • Case one: eg. 3 Columns idendified to be unique/ considered for uniqueness check a. Row level score: Total Unique rows considering only (f1,f2,f3) / total number of records b. Col level score: c. f1_col_score = Total Unique values for f1/ total number of records. d. f2_col_score = Total Unique values for f2/ total number of records. e. f3_col_score = Total Unique values for f2/ total number of records. f. Col Level score = (f1_col_score + f2_col_score + f3_col_score) / 3 g. Uniqueness Score : (Row level score + Col level score) / 2 • Case two: No columns identified to be unique/ considered for uniqueness check a. Row level score: Total Uniqe rows considering all fields (f1,f2,f3, f4, f5) / total b. number of records. c. Col level score = Row level score d. Uniqueness score = Row level score

Metric	Description and data observability score calculation logic
Freshness	**Description:** Measure of how valid is the data as of today. It is measured based on the SLA configured/inferred for date field present in the dataset. If multiple date fields are present, the one closely resembling the transaction date is automatically inferred for timeliness calculations
	Logic: This score is applicable in case there is at least one field with date datatype in the dataset. In case of multiple fields with "date" data type, the one which has latest value close to the current date is considered **Age Score:** Check maximum date from the column and get date difference from current date. Eg. 5 days. If value is less than 2 days (SLA), then age score is 100% else Age Score: SLA(default 2 days)/(day difference between current date and the max date in the field). **Freq Score:** Sort date column from lowest to highest date, compute freq of data (Daily update or weekly or Monthly), eg. If data is updated daily, freq will be 1 day • Freq Score: SLA/(max day gap between any two consecutive dates in a sorted list of dates present in column) Overall Timeliness Score: Mean of Age Score and Freq Score

Metric	Description and data observability score calculation logic
Correctness	**Description:** Measure of how accurate the data observations are. This is calculated as the proportion of outliers in the data. Numeric Fields are considered for the outlier detection as of now. Outliers are values which are 2 standard deviation away from normal
	Logic: • Row Level Score: (Number of rows where numerical fields do not contain outliers)/ (Total number of rows) • Column Level Score: (Number of non outliers values in a field)/(Total number of values in a field). Mean of all such field scores • Overall Score: Mean of Row Level and Column Level Scores
Completeness	**Description:** Measure of the extent to which the mandatory fields are populated in the dataset
	Logic: • Columns which have less than 5% NULL values is considered as mandatory • If more than 5% It is an optional/ non-mandatory. • Completness check is applied on the columns identified as mandatory

Metric	Description and data observability score calculation logic
Validity	**Description:** Measure of the validity of the data in terms of data type, valid data range, valid data values and data format.
	Logic: Different types of logic for validity score calculation • Data Type: Whether the data is of the required data type. • Internally will check dominant datatype and validate data against the dominant datatype • Valid Values: Number of unique values should be less than threshold (20), eg. Gender (M/F), will check to see if column has those values and other values. • Valid range - For numeric fields only, it will compute min and max value and check for to see if values in the attributes fall within that range • Data Format - Based on values in a column, regex pattern is derived from data, derived pattern needs to be 90% to be a dominant pattern. If derived pattern is less than 90% Validity check is skipped. Score is generated based on patterns conforming to dominant pattern

Importance of data observability in finance and accounting business

As finance team is critical for any organization, and they consume data from the other departments to derive growth strategy, financial statements, and plans and budgets. It is important for organizations to have rich data quality and data integrity to avoid any delays in the periodical close process and reporting processes. The key challenge for any finance organization is access to the 'right data at the right time' as they have maximum dependency on other departments. The data observability can be leveraged at two stages to minimize the challenges – one at the transactional or source level and the second at the finance hub data processing.

This would help the business understand the data issues with clarity, whether the source has the data issues or the translation in the middle has the data issues. With this, the business team's productivity would improve during the reconciliation process, and the quality of data would improve for any planning and budgeting process, as the actuals are the basis for any projections.

In addition, the data quality and integrity would help in faster regulatory reporting, and in turn, it helps in meeting deadlines.

LTIMindtree's CARE framework for implementing data observability

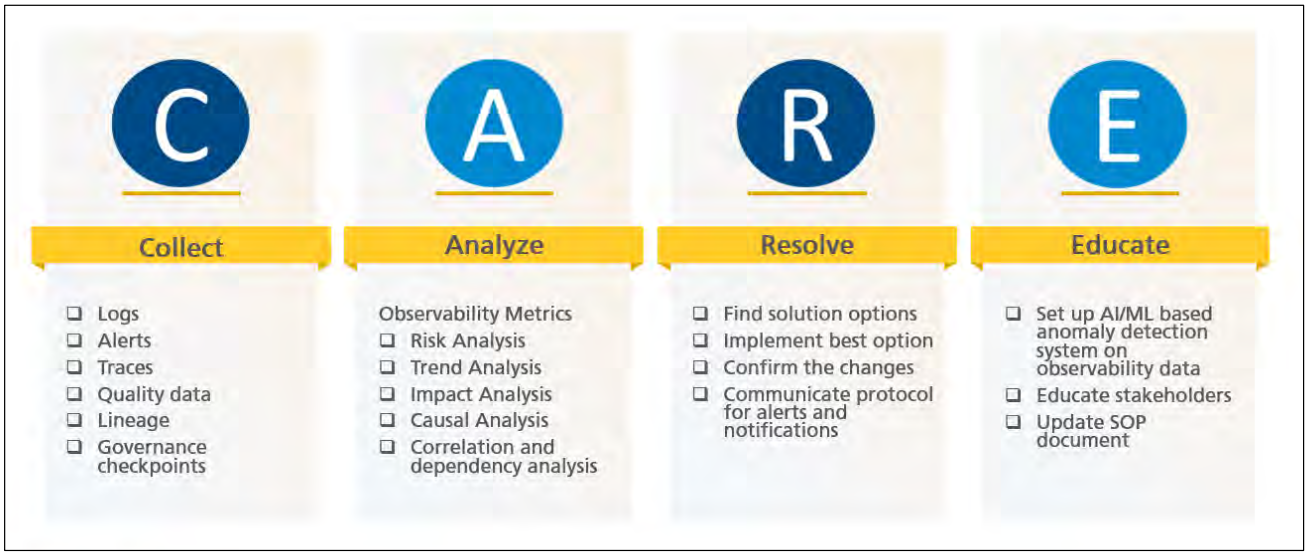

Figure 5: CARE framework for data observability implementation

You can start and mature in your data observability journey using the CARE framework by collecting observability data like logs, traces, and governance checkpoints from different systems used by the organization and analyze them individually initially and draw inferences from them.

Define metrics and KPIs to check observability data in a single view. Later, connect different systems to perform organization-wide impact analysis, where end-to-end data lineage information would be needed. Set up an alert notification mechanism using mail, messaging, WhatsApp, and robo calls.

The next stage is to train the AI/ML model with historical observability data so that an automated agent powered by observability data can detect anomalies faster using historical patterns and alert the stakeholders. Steps taken proactively depending on these alerts may reduce system downtime and improve customer experience.

Beneficiaries of data observability

Data observability will primarily benefit the CxOs of organizations by:

a. Helping make data-driven decisions with increased confidence and accuracy.

b. Ensuring the right data is available at the right time to achieve on-time compliance to minimize the impact on the goodwill of the organization.

c. Improving efficiency by streamlining financial process data integrations and creating financial and non-financial reporting, which would lead to cost reductions.

d. Enhancing stakeholder trust and confidence, leading to stronger investor relations and better business outcomes.

Future of data observability

The future of data observability in financial data lies in leveraging Artificial Intelligence (AI) and Machine Learning (ML) technologies to automate anomaly detection, predictive analytics, and error remonal, enabling real-time insights and proactive risk mitigation. Advanced data observability tools and platforms will evolve to handle larger volumes of data, diverse data sources, and complex data ecosystems, ensuring comprehensive monitoring and analysis. The integration of data observability with emerging technologies such as blockchain, cloud computing, and big data analytics will enhance data security, scalability, and real-time visibility across financial systems.

Key takeaways

As finance is the core pillar of any organization's growth and sustainability, it is important to bring accuracy, reliability, and integrity to the data. It is important that the financial closing processes are completed on time and the planning and budgeting cycles are completed without any delays, as the finance systems are a combination of actuals and forecasting. With the multi-source transactional data and GL data, it is necessary to bring the observability to achieve data-driven decisions at the right time. While business drives the processes and skills required, IT plays an important role in bringing the right investments and advanced technologies with necessary culture changes to unlock the business value from data observability.

References

https://www.gartner.com/en/articles/monetizing-observable-data-will-separate-the-winners-and-losers

About the author

Manojit Saha

Associate Principal – Data Engineering

As a business transformation strategist, Manojit specializes in using digital technology to drive growth, profit, and efficiency for global businesses. He works closely with CxOs to define data and analytics strategies and implement innovative data-driven solutions through automation and digitization. His experience includes leading large global programs, providing thought leadership, strategic consulting, and solution architecting in areas such as data modernization, data engineering, AI/ML, cognitive analytics, visualization, and data storytelling.

Prasanna Ramaswamy

Senior Principal – Consulting

Prasanna is a specialist in Financial Planning and Analysis (FP&A) and group finance solutions and transformation. His experience spans 20 years, with rich knowledge and expertise in consulting, process definition and re-engineering, solution implementations, business process transitions, and building centers of excellence. He extensively worked in implementing, managing, planning, budgeting, consolidating, and reporting solutions. He is passionate about building integrated solutions for financial teams. Prasanna incubated and heading LTIMindtree's Business Finance Performance Management (BFPM) practice.