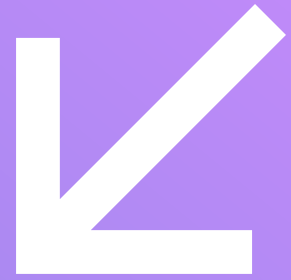


A Guide To Data Observability: How To Make **Your Data** Work For You



Table Of Contents



INTRODUCTION	3
UNDERSTANDING DATA OBSERVABILITY	4
WHY IS DATA OBSERVABILITY IMPORTANT?	9
WHAT ARE THE CHALLENGES OF DATA OBSERVABILITY?	12
DATA OBSERVABILITY FRAMEWORK 7	14
BEST PRACTICES FOR IMPLEMENTING DATA OBSERVABILITY	18
CHOOSING YOUR DATA OBSERVABILITY PLATFORM	20
WHY IS INTEGRATE.IO THE BEST DATA OBSERVABILITY PLATFORM FOR YOU?	23

01

INTRODUCTION

Data observability is becoming an increasingly important part of modern, data-driven organizations as they strive to gain deeper insights into the efficiency of their data ecosystems. Today's data systems and big data trends are continuously evolving, making it difficult to comprehend their intricate behavior and optimize performance - but with the right tools in place, organizations can remain at the forefront of this technological evolution.

In this whitepaper, we will explore the concept of data observability in detail and talk about its importance for your business. We will also discuss the challenges of data observability and lay out a set of best practices to help you navigate them for successfully implementing data observability.

Let us dive deeper into how you can use data observability solutions, like [Integrate.io](https://integrate.io), to make your data work for you.



02

UNDERSTANDING DATA OBSERVABILITY

What Is Data Observability?

Data observability is the process of understanding, diagnosing, and managing data health and movement across multiple systems and organizations. It is an umbrella term that encompasses several concepts associated with identifying, troubleshooting, and resolving data issues in real time.

By measuring and analyzing the inputs and outputs of their data assets, organizations leverage data observability to gain insight into the internal state of their systems. As a result of this analysis, they can find and diagnose data problems and understand the impact of changes made to their systems. Moreover, data observability enables them to predict and prevent potential failures and use observability insights to make data-driven decisions for enhanced system performance and reliability.

The Origin and Evolution of Data Observability

The concept of observability has already been prevalent in software development and operations (DevOps), where it has been used to monitor and understand the performance of complex systems. The activities under observability are designed to tackle the most significant pain point in any software development process; system downtime, where a system is unavailable or not performing as expected.

With the increasing amount of data generated in recent years, organizations have evolved to be more data-driven. To give data its due importance, these organizations started applying the principles and practices of DevOps to data management and analytics, giving rise to DataOps (data operations).

However, as a part of evolving from DevOps to DataOps, the downtime problem developed into a more complex data downtime problem. It was a much bigger problem as the data downtime multiplied with the growing number and complexity of data assets. DataOps needed its own solution to ensure reliable service and this is where 'Data Observability' emerged, transforming DevOps best practices into an invaluable tool in safeguarding modern business operations.

Today, data observability has become an essential aspect of modern software development and operations. With the rise of cloud computing and the adoption of microservices and containerization, data observability has become increasingly crucial for understanding and managing the performance of distributed systems.

What are the Five Pillars of Data Observability?

The field of DevOps has been a significant inspiration behind the concept of data observability. The "three pillars of observability" defined by DevOps are:

1 METRICS

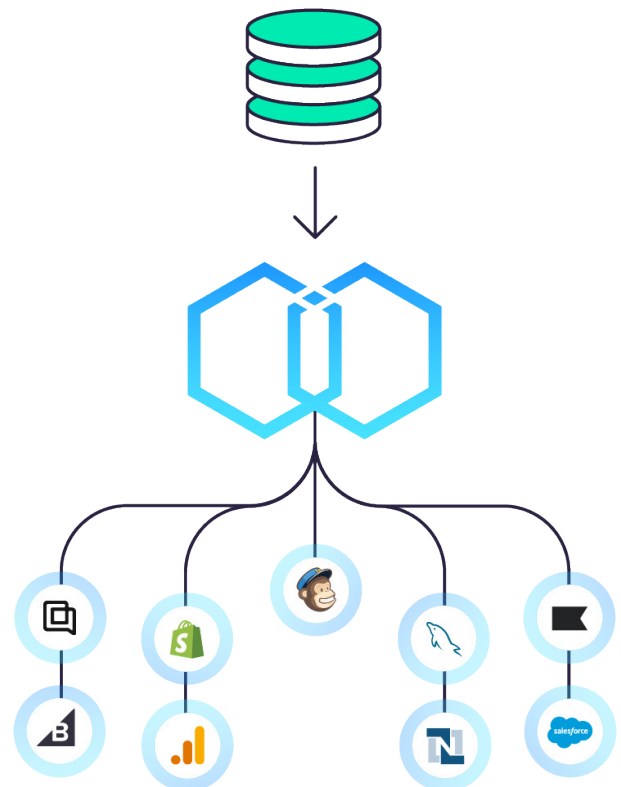
Metrics are numerical measurements of a system's performance, such as CPU usage or network traffic. These provide a high-level overview of how a system is performing and can be used to detect performance bottlenecks or anomalies.

3 LOGS

Logs are records of events that occur within a system, such as error messages or system messages. They provide detailed information about what is happening, answer factual questions about a system's usage and access, and provide root-causes of failures.

2 TRACES

Traces are records of the flow of a request through distributed systems, such as the sequence of microservices that were called to complete a request. These allow teams to understand the end-to-end flow of a request and identify any issues.

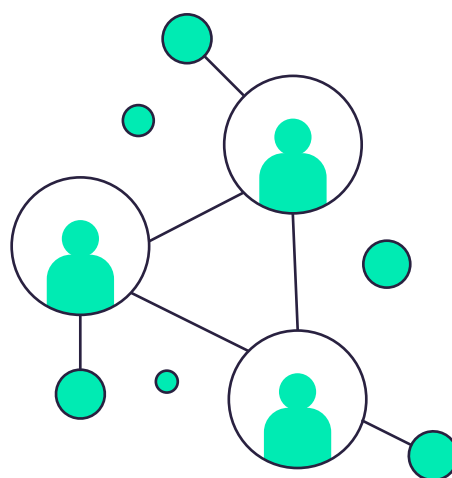


These three pillars serve as the three primary sources of information for DevOps teams to perform root-cause analysis of their IT problems and use troubleshooting to resolve them.

Applying these practices to data observability, data engineering teams have formulated their own "five pillars of observability" as abstract, high-level ideas that capture the gist of data observability that businesses can implement in terms of concrete metrics. Here are the five pillars of data observability:

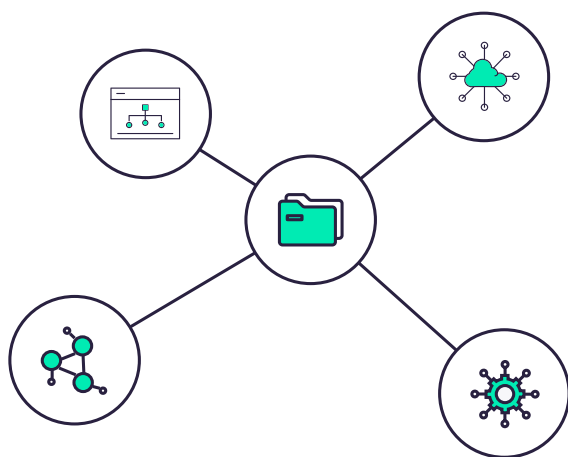
1 FRESHNESS

Informed decision-making heavily relies on having access to the latest, most accurate information. Freshness tracks how new or up-to-date your enterprise data is. The freshness pillar addresses concerns like data recency, frequency of data updates, and more.



2 DISTRIBUTION

The distribution pillar examines the expected value of the data collected by your organization using [data profiling](#). If data doesn't match the expected values, it can indicate an issue with data reliability. Extreme variance in data also indicates accuracy concerns. Strange or anomalous values may signal that the data source is poor quality, unvalidated, or untrustworthy.



3 VOLUME

The volume pillar is associated with the quantity of data your organization has available. Your organization's data volume needs to stay constant or increase. However, this figure can sometimes reduce owing to inaccurate, outdated, or irrelevant data deletion. Monitoring data volume is crucial to assess if your business is effectively intaking data at the expected rate.

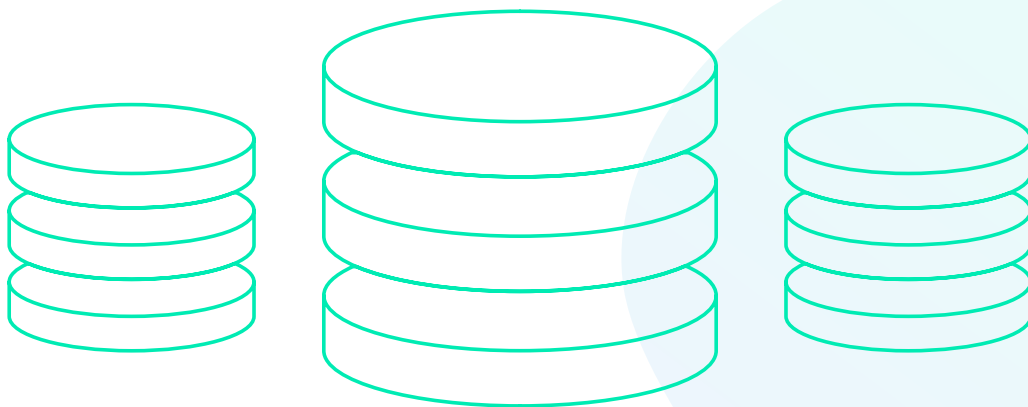
4 SCHEMA

Schema refers to a database or table's abstract design or structure that standardizes how the data is laid out within a repository. Tracking database schemas is helpful to verify that its resident data remains accessible. Changes to a schema are harmful as they can break your database and may indicate accidental errors or even malicious attacks.

5 LINEAGE

And lastly, the fifth pillar of data observability is lineage. lineage is a complete roadmap of your organization's entire data ecosystem.

[Data lineage](#) lays out the flow of information throughout your business, providing a holistic image of your data assets for you to understand and solve critical data problems.



03

WHY IS DATA OBSERVABILITY IMPORTANT?

Data observability is essential because it allows you to understand and monitor the health and performance of your data systems.

In addition to data health monitoring, data observability brings many more benefits to your business. Following are some of the key benefits you can get by adopting a data observability solution:

- Getting the complete picture
- Improved data quality
- Reduced downtime
- Enhanced security
- Increased data efficiency
- Anomaly prevention
- Greater agility
- Cost savings
- Capturing data changes



GETTING THE COMPLETE PICTURE

Making data observability your priority can enable your organization to see better where their data is stored and break down any data silos. It lets your teams and departments understand how to use data effectively in their BI and analytics workloads.

IMPROVED DATA QUALITY

Through real-time data monitoring, you can identify and fix problems that could compromise your [data quality](#), ensuring accurate and reliable data and building better data quality processes over time.

REDUCED DOWNTIME

Data observability serves as a means for businesses to check for data quality issues and limit downtime as much as possible.

ENHANCED SECURITY

A data observability solution supports your data security framework by identifying security breaches or vulnerabilities in your systems, enabling you to take proactive actions and create better security strategies across your data ecosystem.

INCREASED DATA EFFICIENCY

With improved visibility into your data management bottlenecks or inefficiencies, you can leverage data observability solutions to maintain data efficiency throughout the system.

ANOMALY PREVENTION

Besides anomaly detection, data observability provides enough context to enable engineers to resolve the problem and figure out ways to prevent that error from reoccurring.

GREATER AGILITY

Data observability makes it easy for your enterprise to adapt to changing business goals by offering real-time insights into data trends and patterns so you can make informed and meaningful decisions.

COST SAVINGS

The costs associated with finding and fixing data problems can grow as your data changes. Using data observability to get added visibility into your data health allows you to resolve data issues before they escalate and create a more efficient set of systems.

CAPTURING DATA CHANGES

With the continuously evolving business landscape and the resulting data changes, data observability is the only way to keep up with these changes. By providing deep context into every data change, data observability enables you to adapt to the evolving times.

04

WHAT ARE THE CHALLENGES OF DATA OBSERVABILITY?

Implementing a data observability solution in your organization can take a lot of work. It brings several risks and challenges for your organization to tackle.

DATA COMPLEXITY

Complex systems and applications can make understanding the relationships and dependencies between different data sources complex.

DATA SILOS

[Data silos](#) can be a significant bottleneck to aggregating data from different systems and applications, limiting the ability to gain a complete picture of the system.

REAL-TIME MONITORING

The demand for high-speed and near real-time results is not easy to meet. Data observability solutions for large and complex systems are even more challenging to monitor and analyze in real-time.

DATA PRIVACY

Collecting and analyzing large amounts of observability data can raise privacy and regulatory concerns for organizations. It's vital to ensure compliance with relevant laws and regulations and put proper safeguards in place to protect sensitive information.

DATA SECURITY

Data observability tools combine and store large amounts of data in one place for analysis, creating a centralized target for cyber attacks. Ensuring data security and privacy can be challenging, especially when dealing with sensitive information. Moreover, proper access controls are required to mitigate security risks associated with data observability.

DATA ACCURACY

Data observability relies on the accuracy of the data being collected and analyzed, requiring a robust data quality strategy to create a solution with valid insights.

DATA VOLUME AND OVERLOAD

Collecting, processing, and analyzing large amounts of data from various sources can be challenging and time-consuming. Identifying relevant data points for monitoring and troubleshooting can be challenging without the proper skill sets.

BIAS

Data bias is another challenge with gathering data for observability, leading to inaccurate or unfair results.

SCALABILITY

As the volume of data and complexity of systems increase, it can be challenging to scale data observability solutions to meet the needs of the organization.

COST

Data observability solutions can be costly to implement and maintain, and many organizations may not have the budget to invest in them. It's essential to have a well-designed data governance and quality framework to mitigate these risks before implementing any data observability solution.



It's essential to have a well-designed [data governance](#) and quality framework to mitigate these risks before implementing any data observability solution.

05

DATA OBSERVABILITY FRAMEWORK

What is a data observability framework?

A [data observability framework](#) is a systematic layout for implementing data observability in an organization. It is structured to ensure holistic visibility into their systems and apps to track and log their behavior and performance. It also includes the processes, tools, and technologies for data collection from various sources and its storage, processing, analysis, and visualization.

What activities does a data observability framework include?

A typical data observability framework involves the following set of activities to successfully execute observability workflows:

- Data collection
- Data processing
- Data storage
- Data monitoring
- Event Tracking
- Event Logging
- Data analysis
- Root-cause analysis
- Alerting
- Data visualization
- SLA Tracking

DATA COLLECTION

Gathering data from diverse sources such as logs, metrics, traces, and events.

DATA PROCESSING

Processing and transforming data to make it usable for analysis and visualization.

DATA STORAGE

Storing data in a centralized repository for later analysis and visualization.

DATA MONITORING

For providing a continuous operational view of your pipelines to highlight problem areas.

EVENT TRACKING

Setting and tracking priority events.

EVENT LOGGING

Logging a record of an event in a standardized format for faster resolution.

DATA ANALYSIS

Automated analysis to identify patterns, trends, and anomalies and to gain insights into performance applications.

ROOT-CAUSE ANALYSIS

Investigating the root cause of issues to determine the underlying cause and to identify potential solutions.

ALERTING

Generating alerts based on defined thresholds or conditions to quickly identify potential issues.

DATA VISUALIZATION

Visualizing data using dashboards, graphs, and charts to provide a clear and concise representation of the data.

SLA TRACKING

Measuring data quality and pipeline metadata against pre-defined SLAs.

What are the Characteristics of an Efficient Data Observability Framework?

Planning and executing any data framework or process requires meeting certain requirements that determine its effectiveness. Similarly, an efficient data observability solution is characterized by the following features:

- User-friendliness
- Seamless integration
- Data monitoring at rest
- Minimal configuration
- No prior mapping
- Transparent and rich context
- Proactive problem-solving

USER-FRIENDLINESS

Stronger data observability means technical and non-technical users can similarly access and leverage data observability.

SEAMLESS INTEGRATION

An efficient data observability solution fits neatly in your existing data stack, seamlessly connecting to databases, services, and platforms you already use. It enables quick time-to-value and maximum testing coverage. It should not require substantial investments, modifications to existing data pipelines, long hours of manual code writing, or the use of specific languages.

DATA MONITORING AT REST

With an effective data observability strategy, you do not need to constantly move your data around for monitoring, posing security threats. You can monitor information at rest in the location where it is already stored. Such a data observability solution keeps your sensitive information hidden and is highly secure, performant, scalable, compliant, and cost-efficient.

MINIMAL CONFIGURATION

You can get a productive data observability framework up and running with minimal configuration and practically no threshold setting. You do not need to spend resources configuring and maintaining noisy rules. The configuration hassle is mitigated by automated [machine learning](#) models automatically learning your environment and data. Techniques like anomaly detection work behind the scenes so that you can have a holistic view of your data problems and their impact.

NO PRIOR MAPPING

A capable data observability solution takes the load off your organization so you can have comprehensive observability with minimal effort. Instead of requiring you to map the monitoring requirements, it automatically identifies essential resources, dependencies, and invariants, so you can immediately execute your data observability activities.

TRANSPARENT AND RICH CONTEXT

A popular idiom states that the devil is in the details. Data observability solutions take this very seriously by providing rich context so you can rapidly understand and troubleshoot problems and effectively communicate with impacted stakeholders.

PROACTIVE PROBLEM-SOLVING

In addition to providing detailed information about data anomalies, a robust data observability framework ensures you have enough information to fix and prevent issues from happening in the first place. With rich details on data assets, you can responsibly and proactively manage data problems and modifications.

06

BEST PRACTICES FOR IMPLEMENTING DATA OBSERVABILITY

Now that you understand the importance of an efficient data observability framework, it is essential to remember a few best practices before you dive into implementing a data observability framework in your organization:

1 DEFINE OBJECTIVES AND USE CASES

Clearly define what you want to achieve with data observability and what use cases you want to support. It helps you align your activities with your objectives and monitor your observability framework's performance.

2 HAVE PROPER DATA FRAMEWORKS IN PLACE

While data observability covers a wide range of activities closely interconnected with data monitoring, governance, quality, and integrity, it does not replace them. So, to ensure a high-performance data ecosystem, your organization needs to have effective governance, quality, monitoring, and other data frameworks before implementing a data observability framework.

3 PRIORITIZE EVENT TRACKING

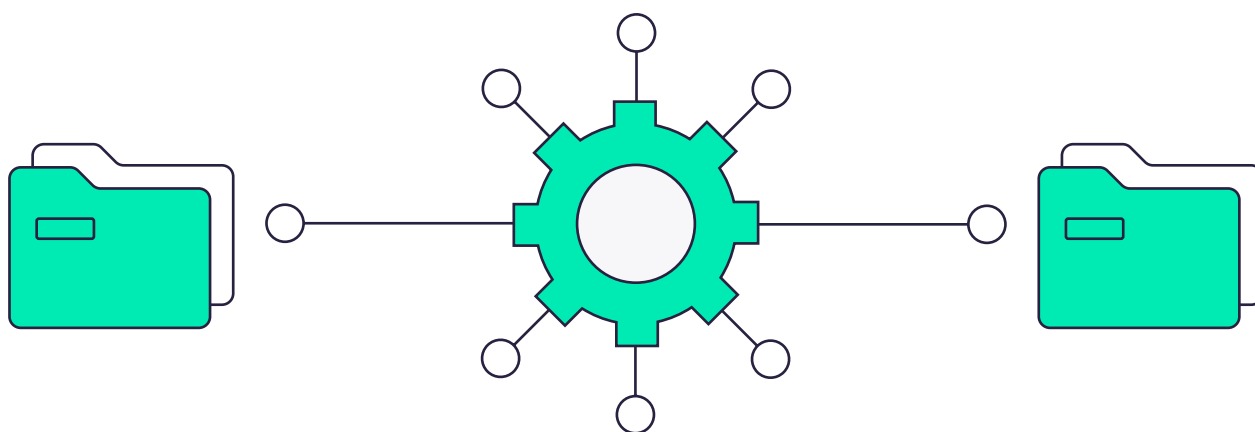
Maintaining traces and logs of complex ETL pipelines and handling millions of data records and hundreds of touchpoints is challenging. To avoid going through millions of lines for an error traceback, it is crucial to focus on and track the information that will provide the most significant value to your business.

4 SET CRITICAL ALERTS ONLY

Just like tracking cannot possibly be turned on for all data records, it can get messy set alerts and notifications for every event. As a best practice, it is advisable to identify and only set alarms for critical events; otherwise, you may disregard the seriousness of problematic alerts in the future.

5 INVEST IN THE RIGHT TOOLS

Creating a successful data observability stack requires investing in the right tools such as log management and monitoring solutions, to support the specific data observability needs of your organization.



6 TRAIN PERSONNEL

In addition to investing in tools and processes, it is important to have educated members in your teams who are trained on your data observability tools and best practices, to ensure they are adequately equipped to use and interpret the data.

07

CHOOSING YOUR DATA OBSERVABILITY PLATFORM

What are the Different Kinds of Data Observability Tools?

There are several types of tools you can use for data observability, including:

1 LOG MANAGEMENT AND ANALYSIS TOOLS

These tools gather, store, and analyze log data from multiple sources to provide impactful performance insights.

3 TRACEABILITY TOOLS

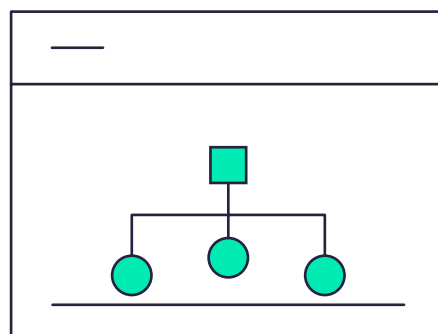
These tools provide visibility into requests and data flow between different data touchpoints.

4 ALERTING TOOLS

These tools enable you to set notifications and alerts for specific events, allowing you to respond quickly to potential issues.

2 MONITORING TOOLS

These tools continuously monitor the health and performance of your data assets to identify potential issues.



5 DASHBOARDING AND VISUALIZATION TOOLS

These tools provide a visual representation of data to help you understand the status and behavior of your data ecosystem.

6 PERFORMANCE ANALYSIS TOOLS

These tools provide detailed insights into performance metrics including response times, resource utilization, and error rates.

You can use these tools individually or in combination to provide a complete picture of data observability for your organization.

When do you need a data observability platform?

Here are a few signs you should look out for that indicate your organization's need for a data observability platform:

1 INCREASING DATA COMPLEXITY

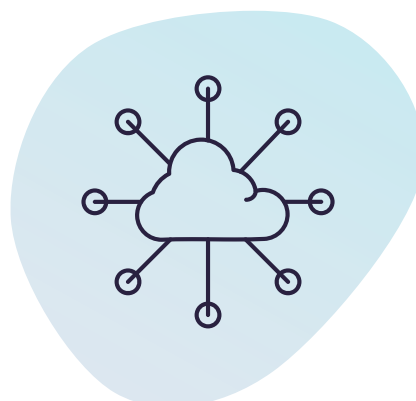
Your data is growing faster than you can keep up with it, affecting your bottom line.

3 SLOW RESOLUTION OF INCIDENTS

Your data anomalies are taking longer to resolve or are not being resolved due to limited visibility.

2 LACK OF ACTIONABLE INSIGHTS

Your existing stack of tools is not providing enough actionable insights, and you need to get more depth into how your business is doing.



4 OVERLOADED DATA STACK

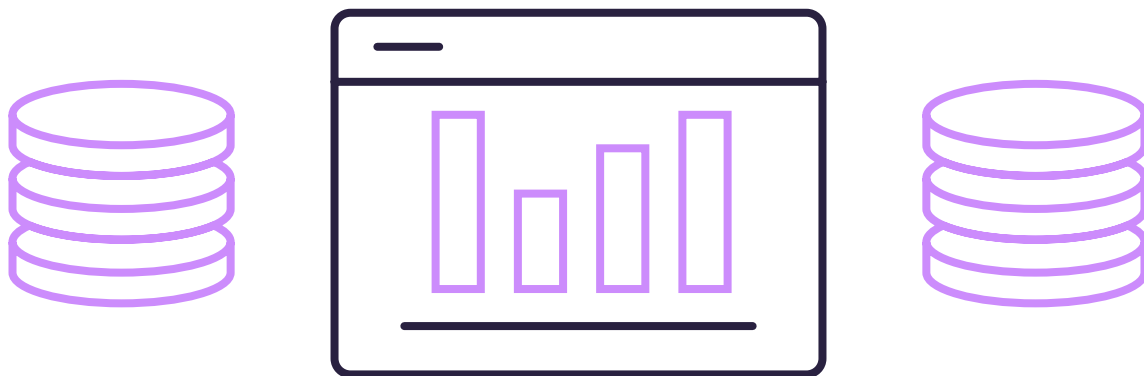
You have a data stack loaded with individual tools for storing and analyzing data, and you need a one-stop solution that you can easily manage.

5 POOR COLLABORATION

You need a collaborative environment where every team can retrieve or interpret the data they need from their systems and communicate actionable insights to stakeholders.

6 INEFFICIENT USE OF RESOURCES

Your organizational resources are not being used effectively, requiring a more efficient approach to monitor and optimize the system's performance.



08

WHY IS INTEGRATE.IO THE BEST DATA OBSERVABILITY PLATFORM FOR YOU?

Integrate.io's data integration platform comes with various data observability solutions that provide broad visibility into your systems and data health. It offers powerful data analytics and transfer capabilities to quickly and efficiently move, query, and analyze data from any source.

Integrate.io lets you keep track of all the data you need to extract the most value from your enterprise data, no matter where it's stored. It enables you to identify problems with your data, check for accuracy, format it correctly, and ensure it complies with your company's data governance guidelines. You can integrate with any data source and pull your data into one place to promptly analyze and make critical business decisions faster.

With Integrate.io, your teams can get instant notifications when a metric goes out of range, so they can promptly tackle the issue. Your DataOps team can specifically benefit from this by quickly detecting any upstream data issues and notifying the relevant team to investigate and resolve the issue.

Some of Integrate.io's key data observability features include:

QUICK AND EASY SETUP

Integrate.io's data observability setup is easy to get up and running to provide you with a complete picture of your data.

COST-EFFECTIVE

As the only observability tool in the market that offers three free alerts on sign-up, Integrate.io is the most cost-effective solution for organizations that are just beginning to implement data observability.

MULTIPLE ALERT TYPES

Integrate.io offers a variety of alerts with options for nine different types of alerts, including Nulls, Cardinality, Median, Variance, Skewness, and Freshness;

REAL-TIME ALERTS

You can set up Integrate to deliver real-time notifications in case of any data issues so you can easily manage and resolve them.

ANALYTICS AND REPORTING

You can analyze trends to identify recurring anomaly patterns in your data sets to resolve possible problems before they arise.

COMPLIANCE

Integrate.io ensures compliance with major data governance frameworks including GDPR, HIPAA, and CCPA.

CHANGE DATA CAPTURE (CDC)

Integrate.io's CDC enables you to easily detect any altered data records and sources after implementing every new integration to help you understand how data changes over time.

REVERSE ETL

Integrate.io makes it easy for you to send data out of a centralized repository and into third-party data systems for more manageable and controlled access and analysis.

Are you ready to let Integrate.io improve your data observability for an enhanced data ecosystem? [Schedule a call with our data experts](#) and start your data observability journey with Integrate.io today.

We hope that you have enjoyed reading this whitepaper.

We hope that you have enjoyed reading this whitepaper. If so, be sure to check out our other Whitepapers related to "[Data Observability: How To Make Your Data Work For You](#)".

Learn more about how Integrate.io can help
make your data work for you!

Chat with an Expert!

