### Introduction

At the heart of every digital transformation effort is the belief that by leveraging new digital technologies, businesses can become more efficient, customer-centric, and cost-effective. They can produce more innovative products, and make better business decisions.

Over the past ten years, many digital transformation efforts have been focused on moving from on-premise infrastructure to the cloud, and from batch-processing small amounts of data to real-time processing of a seemingly infinite stream of data. This data is transformed, manipulated, and ultimately stored in many different places depending on the analytical, product, and budget needs of the company. CIO’s and CTO’s, as a result, are now managing potentially dozens of data ecosystems. To effectively operate in this environment, it is crucial that they have a clear understanding of the various data assets, data pipelines, and business needs that are being supported.

Unfortunately, this is not the reality for most organizations right now. Even as digital transformation efforts have generated a deluge of data, most companies are not in a position to take full advantage of it — they don’t always know where the right data is, and also frequently have to wrestle with wrong/bad data.

This is where data observability can help.

### What is data observability?

Data observability refers to the ability of users, particularly data engineering teams, to monitor and understand the state of their data and its supply chain. This includes being able to identify potential issues or delays in the flow of data, without needing to make significant changes to the system.

More concretely, data observability platforms will generally provide some subset of:

* Monitoring - tracking volume, freshness, quality
* Anomaly detection
* Service Level Agreements (SLAs)
* Data lineage
* Data governance

With these tools, organizations can answer questions such as:

* Is customer data arriving on time?
* Are there any duplicated transactions?
* Is the decrease in average purchase size real or a data issue?
* Will deleting a table from the data warehouse have any impact?

In general, having data observability in place helps prevent data quality issues or at least mitigates their impact on the business.

### Data observability vs. traditional observability

You might be wondering what the difference is between data observability and traditional observability. The main distinction is that while traditional observability focuses on IT applications, data observability specifically addresses the complexities of monitoring data applications and pipelines.

### Data observability vs. monitoring

You might also be wondering about the difference between data observability and monitoring. Monitoring is only one component of data observability that focuses on tracking statistical metrics on your data, and alerting you when those metrics go out of bound. Data observability is a broader concept that also includes things like data lineage (knowing the inputs and outputs of every data pipeline) and metadata.

### When should you implement data observability?

Most teams that are dealing with data quality issues will start by implementing data tests/SQL checks of some kind on the inputs/outputs of their data pipelines. However, these data checks, while a good first step, are ultimately restrictive, difficult to manage, and not scalable.

In general, data observability solutions should be implemented sooner rather than later. A good time is when development is complete, the organization is ready to move into production, and at least one or two business lines are using data analytics and machine learning. By incorporating data observability at this stage, companies can set themselves up for high-quality data from the get-go.

More realistically, companies often turn to data observability after a severe data outage damages the business bottom line or company reputation, or when the company is preparing for an IPO. While your data observability setup may be more complicated, this is also when in investing in it will pay the most high-profile dividends.

To determine whether you are ready for data observability, go through our 20-question checklist [here](https://www.bigeye.com/blog/20-questions-for-data-observability).

### Companies that have already implemented data observability

Data observability is a relatively new independent concept, but versions of it have been implemented in internal efforts at companies like [Uber](https://www.uber.com/blog/monitoring-data-quality-at-scale/), [Netflix](https://www.databricks.com/session_na20/an-approach-to-data-quality-for-netflix-personalization-systems), [Airbnb](https://www.google.com/search?q=airbnb+data+quality&rlz=1C5CHFA_enUS993US994&oq=airbnb+data+quality&aqs=chrome.0.69i59j0i512j0i390l3j69i64j69i60.1991j0j7&sourceid=chrome&ie=UTF-8), and [Lyft](https://www.youtube.com/watch?v=TFcbwOjDpIQ), since the late 2010’s.

Most of these data teams developed some sort of pipeline testing system first, before moving on to developing true data observability tools.

Eventually, smaller companies with lighter technical teams also sounded the alarm for observability capabilities. However, they didn’t have the horsepower to build these solutions in-house. And thus, data observability SaaS solutions were created to fill the gap.

Now, with access to off-the-shelf solutions, major enterprise corporations such as Phone Pay, Walmart, Oracle, and Verisk are investing heavily in data observability.

### Getting executive buy-in for data observability

Now that you understand how data observability can help your digital transformation efforts, how do you go about getting executive buy-in for it at your company? The steps below provide some guidance:

1. Ask questions about the state of the company’s data systems. Examples include:

* "Are you sure that our data is reliable?"
* "How many outages have we suffered due to data issues?"

It may become apparent, after asking some of these questions, that the organization has already been suffering the effects of unreliable data. By highlighting these pain points and linking them to the capabilities of data observability, you can demonstrate value.

1. Cite real-world examples of what can happen when you don’t have visibility into the state of your data systems

For example, at Zillow in 2021, an incorrect data model led to pricing houses out of the market. This can help to illustrate the potential impact on the business of not having proper data observability in place.

1. Address misconceptions or confusion about data observability

Finally, it is important to address any misconceptions or confusion about data observability. An executive might bring up the objection, for example, that the company is already paying for a tool like Datadog. To this, you can clarify that Datadog was built as a standard IT observability platform, to monitor web servers and the like, rather than data assets and data pipelines. Using Datadog to monitor your data pipelines would inevitably involve hack-ish solutions and painful configurations.