### Introduction

Unlike legacy data quality products, which originated in the on-premise era, modern data observability platforms like Bigeye were built from the ground up to handle the complexity and scale of cloud data environments.

In this blog post, we’ll explore how data observability platforms like Bigeye outperform legacy data quality products like Informatica, and why organizations should consider making the switch to a modern data observability platform.

### What is Informatica?

Informatica, founded in 1993, is an all-in-one data integration platform that provides a no-code interface for building ETL pipelines. The platform is designed to help organizations integrate data from various sources into a data warehouse or data lake. Informatica's no-code interface allows non-technical users to build ETL pipelines without writing any code. Informatica is comparable in ethos/paradigm to something like UIPath. Informatica was founded in the pre-cloud era, and many large companies still have on-premise Informatica deployments.

Over the years, Informatica has also expanded into adjacent products in data quality and governance.

### How do Informatica’s data quality products work?

Traditionally, legacy data quality products like Informatica have used rule-based approaches to ensure data quality. These products require users to define and implement rules for validating, cleansing, and transforming data from different sources.

However, these rule-based approaches have several limitations:

* They still fall in the category of writing hundreds or thousands of custom checks for each data source or pipeline
* They do not adapt well to changes in data schemas, formats, or business requirements
* They do not provide comprehensive visibility into data health and performance across the entire data ecosystem
* As a result of everything described above, they require heavy professional services resources to set up and maintain

By contrast, modern data observability products like Bigeye take a much more automated and proactive approach. They enable users to monitor and improve data quality at scale using intelligent methods. In particular, Bigeye offers features like:

#### Autometrics

One of the key features of Bigeye’s data observability product is its autometrics capability. Autometrics are smart recommendations for monitoring coverage based on an analysis of the user’s data, including column type, semantic details, formatting etc. When users connect their database to Bigeye, it automatically indexes their source to generate the catalog and suggest basic autometrics. Customers can then go through and enable the autometrics they want.

#### Autothresholds

When you deploy an autometric that Bigeye recommends, it comes with autothresholds computed from historical data using machine learning models. These autothresholds are dynamic and configurable. You can set them to narrow, normal, or wide. You can also give feedback to the machine learning models on the Issues page if you find that the autothresholds are inaccurate.

#### Anomaly detection

Anomaly detection is the process of identifying and alerting on deviations from expected values or trends in the user’s data quality metrics. ​​Bigeye has built sophisticated anomaly detection algorithms that understand trends and seasonality and that recognize hard-to-detect anomalies such as slow degradation. Intelligent adaptation allows the system to identify and adjust to pattern changes in data, removing the need for manual intervention and facilitating data observability at scale.

Furthermore, Bigeye’s anomaly detection is continually refined through reinforcement learning and anomaly exclusion, enabling the system to learn from user feedback and fine-tune detection and alerting capabilities. Finally, Bigeye provides a user-friendly interface that aids in root cause and impact analysis, as well as related issue identification, to help engineers resolve detected anomalies effectively.

#### Metadata metrics

Bigeye offers special, Metadata metrics immediately upon connecting your data warehouse. Metadata metrics encompass metrics like data freshness and volume, that generally indicate the success or failure of a data pipeline. These metrics primarily focus on whether a table has been updated and/or accessed. The table below presents some examples:

| **Metadata Metric Name** | **API Name** | **Description** |
| --- | --- | --- |
| Hours since last load | HOURS\_SINCE\_LAST\_LOAD | The time elapsed (in hours) since a table was last modified with an INSERT, COPY, or MERGE operation. Recommended as an automatic metric for each table. |
| Rows inserted | ROWS\_INSERTED | The total number of rows added to the table through INSERT, COPY, or MERGE statements in the last 24 hours. Recommended as an automatic metric for each table. |
| Read queries | COUNT\_READ\_QUERIES | The total number of SELECT queries executed on a table in the last 24 hours. Recommended as an automatic metric for each table. |

Metadata metrics are a crucial component of Bigeye's T-shaped monitoring approach, which advises tracking basic metrics across all data while implementing more in-depth monitoring for the most critical datasets, such as those used in financial planning, machine learning models, and high-level executive dashboards.

### Conclusion

In conclusion, rule-based data quality solutions like Informatica (or even more modern tools like DBT tests) require users to manually define each metric and their associated thresholds. These rules can be complex and time-consuming to maintain and update as the data changes over time. Additionally, a lack of sophisticated anomaly detection algorithms in these products means that anomalies detected might be false positives or not actionable. By comparison, Bigeye provides instant coverage for the entire data warehouse from the moment customers connect, with the option to go deeper with little additional effort.

If you want to learn more about how Bigeye can help you achieve better data quality at scale, sign up here.