

Coverage Initiation: Bigeye doesn't blink on data observability, bringing in fresh funding

Analysts - Paige Bartley

Publication date: Wednesday, September 29 2021

Introduction

Data quality has always been a challenge for organizations, more so today than ever. More data sources, more data formats, more data integration methods – the highly heterogeneous enterprise data environment does not lend itself well to validating data quality against static sources of information that may be only periodically updated. The data landscape is now more dynamic, with examples such as IoT data breaking the mold for traditional data quality and validation efforts. According to 451 Research's Voice of the Enterprise: Internet of Things, Budget & Outlook 2021 survey, 39% of respondents report that 'data management/quality tools' are among the most important IoT storage product attributes. Clearly, newer and higher-velocity sources of data are an underserved area in data quality, although data quality as a whole remains a challenge, with many historical problems still persistent.

This need points upstream in the data pipeline, suggesting that organizations must continuously monitor their living data environments for potential data quality anomalies before those issues manifest into functional downstream problems for lines of business (LOBs), which depend on data for insight. 'Data observability' specialist Bigeye (formerly known as Toro) is treating data quality more like a continuous engineering and monitoring challenge rather than an ad hoc remediation challenge. Infused with a fresh \$45m series B round of funding, the company's approach may be validated.

The 451 Take

Bigeye may be 'swimming' against some of today's pervasive data quality current, but it certainly has meaningful ambition. The vendor – named more after the powerful fish than the anatomical feature – is charting a course to redefine what data quality really means. Not only does data need to be

observed in its native environment to catch potential anomalies before they become quality problems downstream, but Bigeye posits that data quality itself also needs to additionally be an internal SLA between organizational data engineering teams and LOBs that ultimately consume data. This philosophy breaks the mold of most cyclical enterprise data quality efforts, but Bigeye thinks it can establish a new norm. Fresh funding may help it.

The effort is not without challenges. While Bigeye may potentially excel at data observation and monitoring techniques, it is not an automated data quality remediation product. Pinpointing problems much further upstream in the data pipeline naturally lessens the impetus to go and 'fix' data in a labor-intensive way, but some of those challenges could continue to exist. In this sense, the vendor's offering has ample room to expand its technical functionality.

Context

Bigeye is the brainchild of cofounders Kyle Kirwan (CEO) and Egor Gryaznov (CTO). The company evolved out of data engineering and science pain points that were experienced in some of the early days at Uber, the well-known ridesharing firm where both cofounders met and initially collaborated. The two worked on multiple teams together in 2014-2015, with Kirwan more focused on data science-oriented roles and Gryaznov focused more on data pipelining and engineering. With money riding on whether the data was correct at the company, there was also a simultaneous explosion in workers setting up their own pipelines – at Uber during this time, anyone with SQL access was able to modify pipelines. There had to be additional investment in an operations layer, as the vendor was being impeded by the weight of its own analytics and data activity.

This is where the cofounders came up with the concept of a software layer or platform that would focus on data reliability engineering. The goal was to treat data quality like an engineering problem, as an ongoing engineering process analogous to DevOps function within the business. The initial focus of the product was the concept of observability, a concept borrowed (sure enough) from the engineering world, but applied to data itself. Arguably, observability of data flows and systems is the first step to enabling data reliability engineering to ensure that data is continuously fit for use within an organization. Consistent monitoring is critical to catch potential problems early.

While data observability as a concept is not entirely unique to Bigeye, one potential differentiator is the company's methodology regarding how data quality should be ensured on an ongoing basis. Data quality scoring mechanisms, often (arguably) rudimentary, have commonly been built into many data management platforms and products over the past few years, with varying results. Bigeye is making the case that algorithmic scoring, particularly against some sort of invisible standard, is essentially meaningless. Instead, it is advocating for an internal SLA model between data engineering teams and data consumers within individual organizations – it believes data quality should be an ongoing delivery and contract arrangement rather than an arbitrary score.

To date, the vendor has raised a total of \$66m in funding across seed, series A and series B rounds. The most recent of these capital infusions occurred in September, when a \$45m series B was announced that was led by Coatue with participation from existing investors Sequoia Capital and Costanoa Ventures. Bigeye primarily intends to use the funds to scale the team and build on existing product leadership.

Notable current customers include Clubhouse, Instacart, Mindstrong, Recharge and Udacity. The company has approximately 20+ employees – although that is projected to grow soon with the new infusion of capital. Bigeye is based in San Francisco. However, its team is generally distributed via a work-from-home model.

Technology

The vendor's technology platform is intended to be tightly integrated into the enterprise analytics workflow, plugging directly into the enterprise data warehouse (EDW) environment, most often in the cloud. This can include product examples such as Redshift or Snowflake, or even open source options like Presto. The platform uses a JDBC connection and read-only account to connect to data sources, enabling collection of data health metrics.

One of Bigeye's core aims is to instrument the data warehouse with living metrics for the data, including measures such as standard deviations and duplicates. This capability is underpinned with automation via the Autometrics feature, and automation can be tuned to thresholds as determined by the customer organization. In essence, unique algorithms can be used to determine individual metrics for data quality. So-called Autothresholds can – by default – adjust themselves over time by analyzing and learning the behavior of an organization's data health tracking metrics. These can also be manually overridden, with custom flat or seasonal thresholds to account for specific trends. Autothresholds can additionally be set to different sensitivity levels and react to user feedback, so that the supporting models don't 'learn' the wrong things.

Alerting is an important function of the company's platform, and the alerting capacity is designed to embed directly into the existing preferred workflow of the organization through other collaborative tools. Data quality and anomaly alerts can go out to qualified enterprise users such as data engineers via platforms like Slack, email platforms such as Microsoft Outlook, PagerDuty, or other communications options through webhooks. From this integrated alerting system, users can log directly into Bigeye whenever appropriate and perform root cause analysis to get to the bottom of data quality issues, thus paving the way for meaningful remediation. As a company, Bigeye is already underway in prototyping ways for its platform to make API calls into other parts of infrastructure so that data quality issues can eventually be resolved even before escalating to a human operator.

However, specific data quality anomalies are typically quite unique to an organization's own data environment, and thus are difficult to simply automate away in terms of remediation. To address this, Bigeye is helping organizations move forward with their remediation efforts by using an Apache Airflow operator to make enabling Autometrics and Autothresholds part of the data pipeline authoring process.

Competition

Apples-to-apples comparisons are always difficult in the data management market, but Bigeye likely finds its closest companions to be players such as Datafold, [Lightup](#), [Monte Carlo Data](#) and Soda Data. All of these providers, while arguably each differentiated in their own way, have focused on upstream detection of data quality anomalies early in the data quality pipeline so that potential problems can be remediated as soon as possible before they spiral into more problematic issues with downstream data-dependent insight initiatives. The concept of data observability is not exclusive to Bigeye, and many of these other firms also lean on the data observability term to describe functionality. They collectively can be considered part of a new vanguard of data quality vendors focused more on data engineering and monitoring rather than data quality initiatives.

The 'traditional' data quality market cannot be ignored here, although it may be different in its architectural approach. Upfront data profiling, in particular, was one such approach by traditional data quality providers to catch quality issues further upstream. Bigeye could have some nominal competitive overlap with that traditional market, as it is technically assisting organizations in addressing some of the same underlying pain points with data quality. Longtime major players in the data quality space include Experian Data Quality, IBM, Informatica, Oracle, Precisely (including both

Trillium and Infogix assets) and Talend. Small and midsize firms that specialize in data quality or have data quality capabilities include Ataccama, Datactics and Syniti.

It would be remiss here to dismiss enterprise efforts to build custom, homegrown systems for data monitoring and quality. Bigeye itself was spawned, so to speak, from experience at Uber, a company that has notoriously built its own data systems from the ground up rather than relying on vendor offerings. For some of the largest and most data-intensive accounts, the DIY mentality might very well still represent the biggest rivalry to the packaged Bigeye platform.

SWOT Analysis

Strengths	Weaknesses
Low data quality isn't just an artifact of poor data management, it is the symptom of an entire living data lifecycle. Bigeye is looking to 'fix' data quality as far upstream in the data pipeline as it technically can, looking directly at infrastructure. This approach is technically robust, and hooks into the customer organization's ability to take action to remediate identified problems via preferred tools.	The vendor is not necessarily unique in its promotion of the data observability concept. While data observability is entirely valid in technological theory, others are also grasping at this marketing term. Bigeye will potentially need to further differentiate itself in this developing market of new data quality-affiliated entrants with additional emphasis on automation (which it seems to be addressing) or other IP-bound capabilities.
Opportunities	Threats
Traditional data quality methods are growing stale to some extent, particularly when it comes to in-motion or streaming data sources in scenarios such as IoT data collection. The ability to monitor these data sources is a growing opportunity for companies like Bigeye that nominally participate in the data quality space via other angles or new architecture.	No matter how feisty or well-funded, a startup in an active enterprise tech market is a potential target for acquisition. In data quality, this might be especially true. The legacy or traditional providers in this space, which have long been focused on data quality remediation, may be eager to make a strategic move to expand their capabilities. A pickup might nullify some of the neutrality that the Bigeye brand can currently offer.

Source: 451 Research