Data reliability is about ensuring that your data is accurate, consistent and trustworthy at all times. It's essential for any data-driven organization that wants to make informed decisions, deliver value to customers and stakeholders, and comply with regulations.

However, many organizations struggle with data reliability issues due to various factors such as complex and dynamic data pipelines, lack of visibility and governance, human errors and biases, and insufficient tools and processes.

How do you know if your organization needs to invest in data reliability? Here are nine common signs that indicate it's time to take action:

1. Nobody trusts your internal analytics/dashboards

One of the most telling signs that your data is unreliable is when nobody trusts your internal analytics or dashboards. When executives see a number, think it's "too low," and immediately attribute it to bad data, this 1) reduces the likelihood of them making data-driven decisions and 2) occasionally might blind them to what the data is actually telling them. This lack of trust can be detrimental to decision-making and lead to missed opportunities.

1. Your engineers and data scientists ignore most of the data alerts they get

If your engineers and data scientists receive many alerts about potential data issues, but they ignore most of them because they're false positives or not urgent enough, then you have an alert fatigue problem. Your data alerts should be meaningful and timely, so that you can quickly detect and resolve any anomalies or errors in your data pipeline. If they're not, then you're risking missing critical issues that could impact your business outcomes.

1. You had an incident impact your customer-facing ML models

Customer-facing incidents precipitated by data issues are one of the most painful (yet common) ways that an organization realizes that they need to invest in data reliability. Especially now that many companies are running ML models in production, giving real-time recommendations, the stakes are higher than ever for the data that underlies the models’ predictions. Your ML models should produce accurate and consistent predictions or recommendations that enhance customer experience and satisfaction. If they don't, then you're exposing yourself to potential losses or damages due to poor decisions or actions based on faulty data.

For example, imagine if your ML model for setting customer credit limits used a data source that went to zero for several weeks in a row due to a bug in the pipeline. This could result in drastically lowering customer credit limits without any valid reason, causing rejected purchases and unhappy customers.

1. Your data quality initiatives keep getting bogged down, over budget, and failing

Data quality initiatives are often launched to improve data reliability, but they can sometimes get bogged down, over budget, or fail altogether. Some common reasons why include a lack of clear objectives and a lack of alignment between the different stakeholders. If your data quality initiatives have felt nebulous and ineffectual, it’s important to tie the investment to measurable metrics, like NPS scores, and to business outcomes.

1. You have a huge number of duplicate tables

If you have a huge number of duplicate tables, it's generally because people don't know where to find data, so they keep on creating new tables. This can lead to inconsistencies and inaccuracies across key metrics, which can have a ripple effect throughout the organization. Investing in data reliability can help establish a single source of truth for your data, which can reduce confusion and errors.

1. PMs are unable to answer simple questions that help them make product decisions in a reasonable amount of time

To see whether you need to invest in data reliability, you can have your PMs run a simple test: have a newly onboarded PM answer some simple analytics questions, such as how many users are using a certain feature, how often they use it, or what impact it has on retention or revenue. If they can’t answer the question in a reasonable amount of time, it’s a clear sign that there are data reliability issues in the organization. Your product managers should be able to leverage reliable and timely data insights to make informed and effective product decisions. If they can't, then they will miss opportunities to innovate, optimize or pivot their products based on customer feedback or market trends.

1. It's someone's job to "babysit" the data pipeline, or to manually debug data discrepancies

If it's someone's job to "babysit" the data pipeline or to manually debug data discrepancies, it's a sure sign that your data pipeline is not reliable. Not only does this take up valuable time and resources that might be deployed towards other data engineering projects, it’s unlikely the babysitter will be able to debug every single issue, which means that data issues inevitably get dropped. By investing in data reliability, organizations bring much more rigor to the babysitting process. Rather than reacting to data issues, you can proactively detect and resolve them. Rather than debugging data issues one by one, you can correlate them.

1. You deliberately schedule the data pipeline to run on Fridays so that engineers have the weekend to debug issues

Organizations have been known to schedule data pipeline runs for Fridays, so that errors may be debugged over the weekend. Like having someone babysit the data pipeline, this is a coping mechanism for a lack of data reliability. In an ideal world, your data should be ready for consumption at any time, so that you can deliver fresh and accurate data to your users and stakeholders on demand. If it isn’t, you're compromising your data quality and timeliness, and putting unnecessary pressure on your engineers.

1. You are planning an IPO

Once your company goes public, you will be required to file accurate and auditable data reports

on a regular basis to meet various regulatory standards and expectations. If your data is unreliable or inconsistent, then you will face legal risks and reputational damages due to potential errors or misstatements in your filings.

### How Bigeye can help

If any of these signs resonate with you, then it's time to invest in data reliability with Bigeye. Bigeye is a data observability platform that helps you monitor, measure, and improve your data reliability across your entire data stack.

With Bigeye, you can:

- Automatically discover and catalog all your data sources

- Track and validate key metrics for data quality, freshness, distribution, lineage, and more

- Detect and alert on any anomalies or errors in your data pipeline

- Drill down into root causes and remediation actions for any data issue

- Generate comprehensive and customizable reports on your data reliability status and trends

Ready to take control of your data reliability?

Start a free trial with Bigeye today!