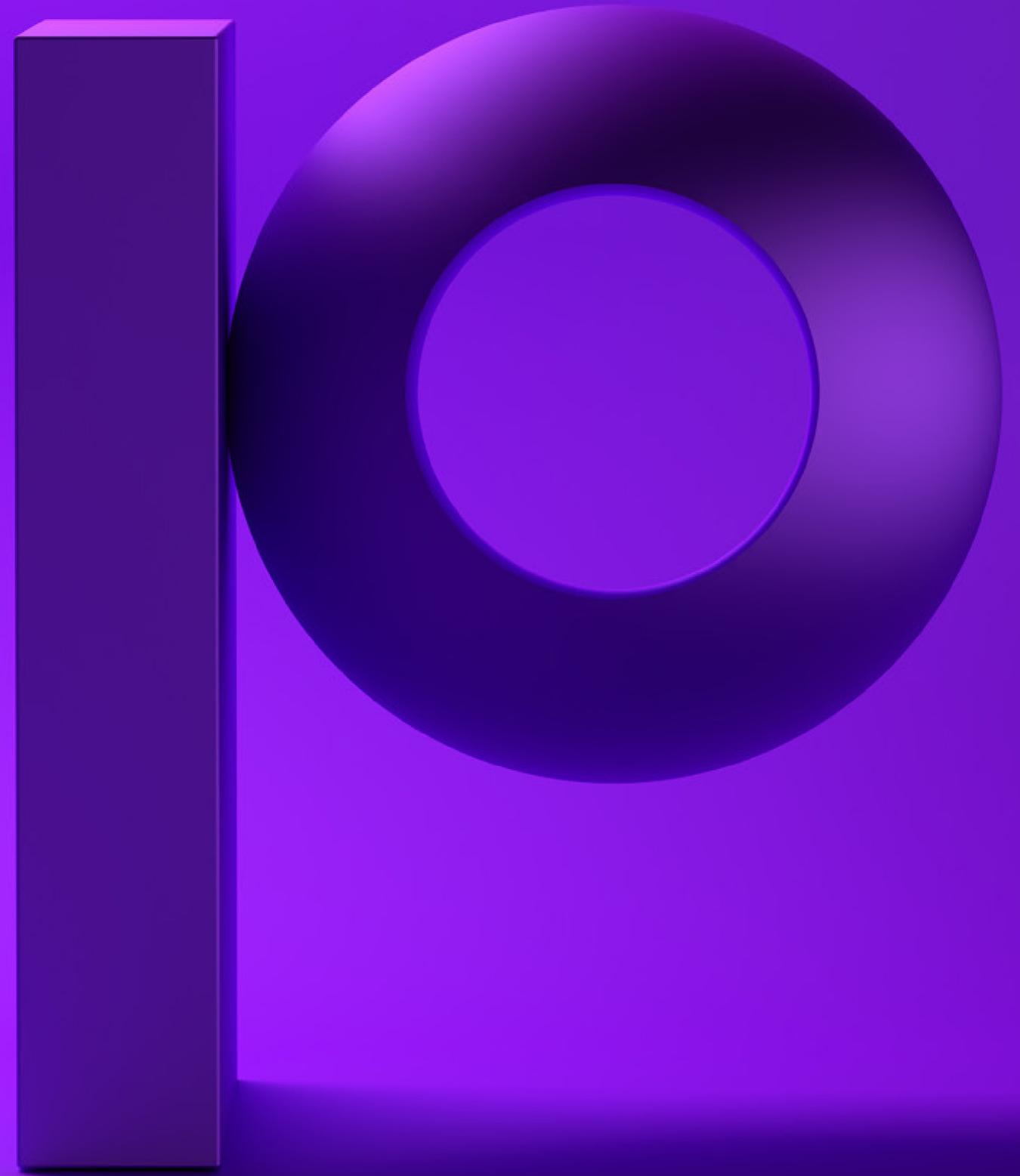


precisely

Seven Metrics to Assess Your Data Quality in Collibra

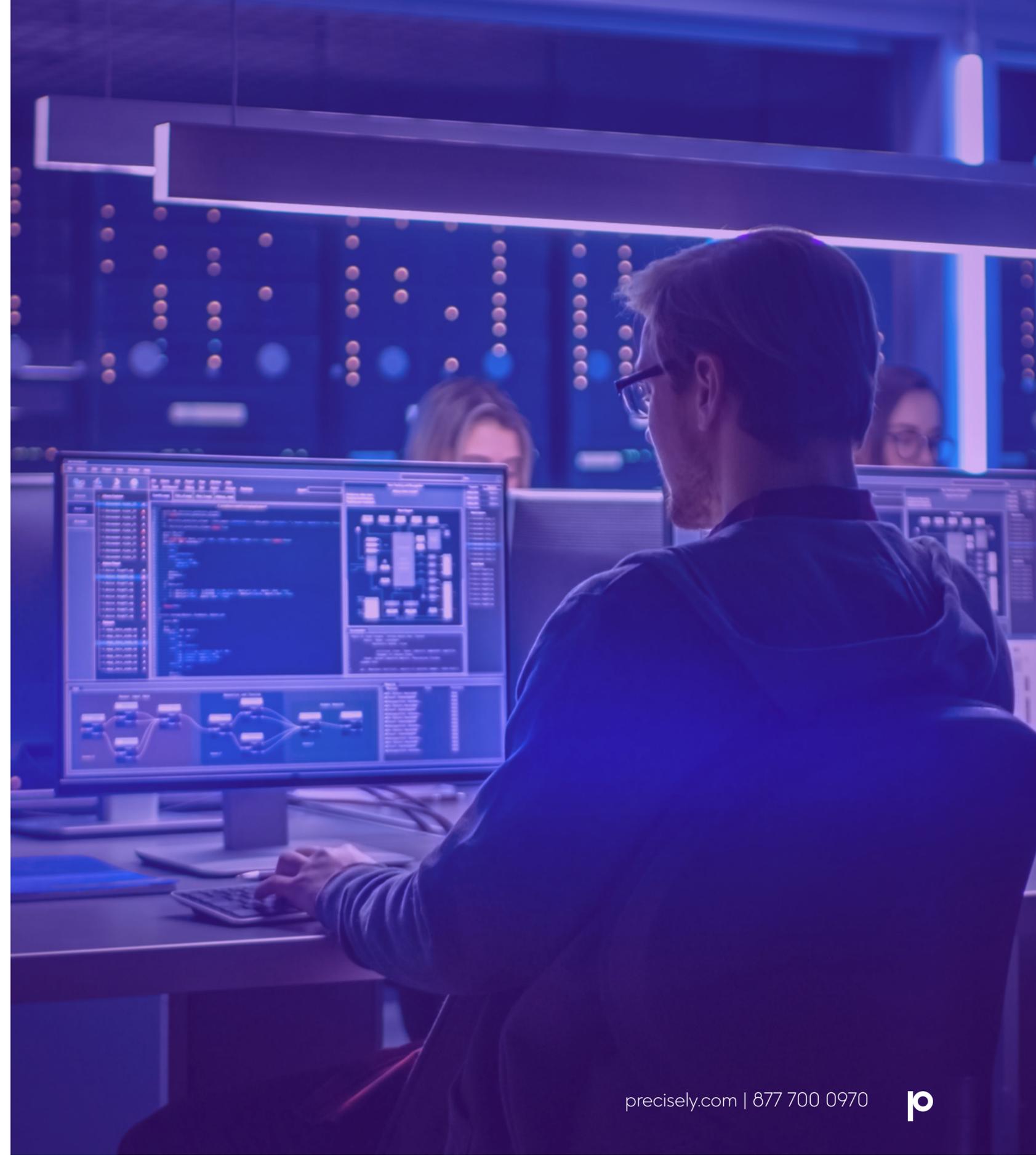


Introduction

Your business depends on accurate data. Inaccurate, incomplete, inconsistent data diminishes the quality of customer experiences, hinders operational efficiency, and threatens regulatory compliance, ultimately exposing your organization to unnecessary risk, instead of giving you the information you need.

Data governance initiatives seek to solve these problems, and to provide the business with trusted, high quality data that will boost marketing effectiveness, customer satisfaction, and ultimately revenue. Data governance tools like Collibra Data Governance Center provide a broad set of capabilities to identify and manage datasets. Sustainable data governance requires a solid foundation of quality data. It requires the right people, the right processes, and the right technology to turn raw, untamed data into valuable business insights.

Data quality refers to the ability of a set of data to serve an intended purpose. Low-quality data cannot be used effectively to do the thing with it that you wish to do. Data quality and data governance share a 'symbiotic relationship'. Data governance needs appropriate data quality tools to not only clean the raw data, but to illustrate data errors, peculiarities and issues, in order to help compile the best standards and monitor the data quality against policies for critical data elements over time. Precisely partnered with Collibra to make industry-leading data validation and data quality monitoring capabilities an integrated component of the Collibra Data Governance Center.



Assessing Data Quality

There are lots of good strategies that you can use to improve the quality of your data and build data best practices into your company's DNA. Although the technical dimensions of data quality control are usually addressed by engineers, there should be a plan for enforcing best practices related to data quality throughout the organization.

After all, virtually every employee comes into contact with data in one form or another these days. Data quality is everyone's responsibility. Assessing data quality on an ongoing basis is necessary to know how well the organization is doing at maximizing data quality. Otherwise, you'll be investing time and money in a data quality strategy that may or may not be paying off.

To measure data quality – and track the effectiveness of data quality improvement efforts – you need, well, data. What does data quality assessment look like in practice? There are a variety of data and metrics that organizations can use to measure data quality. We'll review a few of them here.



7 Metrics to Measure Data Quality

The most obvious and direct measure of data quality is the rate at which your data analytics processes are successful. Success can be measured both in terms of technical errors during analytics operations, as well as in the more general sense of failure to achieve meaningful insight from a dataset even if there were no technical hiccups during analysis.

The fewer data quality problems you have to start with, the faster you can turn your data into value. The main purpose of a data quality plan is to enable effective data analytics, so fewer analytics failures mean you are doing a good job on the data quality front.

Below are seven metrics to help you get started on your data quality plan:

Metric	Definition	How to calculate
Ratio of Data to Errors	How many errors do you have relative to the size of your data size?	Divide the total number of errors by the total number of items.
Number of Empty Values	Empty values indicates information is missing from a set.	Count the number of fields that are empty within a data set.
Data Transformation Error Rates	How many errors arise as you convert information into a different format?	How often does data fail to convert successfully?
Amounts of Dark Data	How much information is usable due to data quality problems?	Look at how much of your data has data quality problems.
Email Bounce Rates	What percentage of recipients didn't receive your email because it went to the wrong address?	Divide the total number of emails that bounced by the total number of emails sent, then multiply by 100.
Data Storage Costs	How much does it cost to store your data?	What is your data storage provider charging you to store information?
Data Time-to-Value	How long does it take for your firm to get value from its information?	Decide what "value" means to your firm, then measure how long it takes to achieve that value.



1. The ratio of data to errors

This is the most obvious type of data quality metric. It allows you to track how the number of known errors – such as missing, incomplete or redundant entries – within a data set corresponds to the size of the data set. If you find fewer errors while the size of your data stays the same or grows, you know that your data quality is improving.

2. Number of empty values

Empty values – which usually indicate that information was missing or recorded in the wrong field — within a data set are an easy way to track this type of data quality problem. You can quantify how many empty fields you have within a data set, then monitor how the number changes over time.

3. Data transformation error rates

Problems with data transformation – that is, the process of taking data that is stored in one format and converting it to a different format – are often a sign of data quality problems. Your data transformation tools will struggle to work effectively with data that they encounter in unexpected formats, or that they cannot interpret because it lacks a consistent structure. By measuring the number of data transformation operations that fail (or take unacceptably long to complete) you can gain insight into the overall quality of your data.

4. Amounts of dark data

Dark data is data that can't be used effectively, often because of data quality problems. The more dark data you have, the more data quality problems you probably have.

5. Email bounce rates

If you're running a marketing campaign, poor data quality is one of the most common causes of email bounces. They happen because errors, missing data or outdated data cause you to send emails to the wrong addresses.

6. Data storage costs

Are your data storage costs rising while the amount of data that you actually use stays the same? This is another possible sign of data quality issues. If you are storing data without using it, it could be because the data has quality problems. If, conversely, your storage costs decline while your data operations stay the same or grow, you're likely improving the data quality front.

7. Data time-to-value

Calculating how long it takes your team to derive results from a given data set is another way to measure data quality. While a number of factors (such as how automated your data transformation tools are) affect data time-to-value, data quality problems are one common hiccup that slows efforts to derive valuable information from data.

Summary

The metrics that make the most sense for you to measure will depend upon the specific needs of your organization, of course. These are just guidelines for measuring data quality. Precisely offers [data quality products](#) that seamlessly integrate with Collibra for a complete, closed-loop data governance solution. Build your data quality rules in Collibra, and they are automatically passed to Precisely Trillium for data quality processing. The data quality results and metrics are then passed back to Collibra – allowing data stewards and business users to see the health of the data right within their Collibra dashboard. The most important thing is to have some kind of data quality assessment plan in place, whatever its details may be.

The importance of data quality, and the amount of data you have to process will only increase with time at most organizations. Continually improving your ability to maintain data quality will help keep you prepared for the data analytics requirements of the future.





Precisely is a global leader in data integrity, providing accuracy and consistency in data for 12,000 customers in more than 100 countries, including 90 percent of the Fortune 100. Precisely enables companies to integrate, verify, locate, and enrich their data to power better business decisions. To learn more, visit www.precisely.com.

www.precisely.com