Collibra

Data Quality and Observability for data warehouse, lakes and lakehouses

Your data warehouse, lake or lakehouse serves as a primary repository, fueling traditional machine learning, generative AI, AI agents, business intelligence, regulatory reporting, and data products. To maximize the value of this data, it's essential to ensure its fitness for as many use cases as possible. The reliability and trustworthiness of your AI and analytics depend heavily on high-quality data in your warehouse, lake or lakehouse.

Deploy data quality observability

Forty two percent of organizations have not deployed data quality observability. Proactive monitoring, validation, and notification can help you find and address data quality issues before they become business issues.

1

30%

of data in the data lake or lakehouse is highly trustworthy and production ready.

Deploy data pipeline observability

Forty five percent of organizations have not deployed data pipeline observability. Proactive monitoring, validation, and notification can help you find and address data pipeline issues before they become business issues.

2

27%

of data pipeline outputs are highly trustworthy and production ready.

Certify data sets and reports

Label certified data sets and the reports that use those data sets. Also make data quality scores and lineage viewable from within your data catalog to create confidence in data.

3

45%

of BI reports and dashboards are highly trustworthy and production ready.

Establish governance policies and processes

Formal policies and processes for managing data quality enable faster and more transparent resolution of data issues. They also help the different stakeholders work together more effectively and efficiently on issue management.

4

18%

of organizations have formalized and consistent policies.

Looking for more best practices for data, pipeline, and AI model observability in traditional ML and GenAI? Read this BARC report.